

Controlling Health Care Costs by Direct Charges to Patients: Snare or Delusion?

Ontario
Economic
Council

CA 26N
EC 17
-79 P10

M.L. Barer, R.G. Evans and G.L. Stoddart

Occasional Paper 10

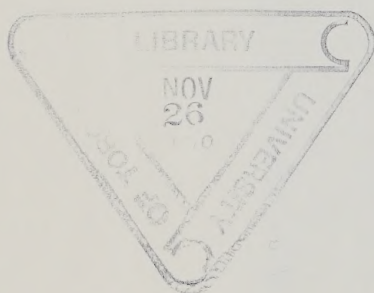
Ontario Economic Council©
81 Wellesley Street East
Toronto, Ontario
M4Y 1H6

1979

Printed in Canada

ISBN 0 7743 3895 4

ISSN 0703-5063



This study reflects the views of the authors and not necessarily those of the Ontario Economic Council

Contents


Figures v

Tables vi

Preface vii

Acknowledgments xi

- 1 Introduction: background and objectives 1
 - Historical perspective: costs and utilization 4
 - Historical perspective: consumer participation 10
 - On current snares and delusions 13
 - 2 Health system objectives and the role of health insurance 19
 - Health insurance objectives 20
 - Health care delivery system objectives 22
 - 3 Charges uniform across providers 25
 - Coinsurance 27
 - Deductibles 51
 - Per-service charges 56
 - Income- and income tax-linked options 60
 - 4 Provider-determined differential charges 73
 - Major-risk medical insurance 76
 - Extra-billing by physicians 82
 - Service repackaging 90
 - 5 Non-provider determined differential charges 97
 - Selective deinsurance 99
 - Parallel systems 104
 - 6 On avoiding future snares: summary and policy conclusions 109
- References 117



Digitized by the Internet Archive
in 2024 with funding from
University of Toronto

<https://archive.org/details/39221624020035>

Figures

- 1 Methods of increasing patient participation in payment for health care through direct charges 17
- 2 Methods of increasing patient participation in payment for health care: charges uniform across providers 28
- 3 Comparison of consumer participation schemes in relation to out-of-pocket expenses of patients 53
- 4 Sample plans illustrating income class-linked patient cost sharing 62
- 5 Consumer risk-bearing in Ontario Economic Council schemes 65
- 6 Consumer risk-bearing in income-tax-linked schemes 69
- 7 Methods of increasing patient participation in payment for health care: provider-determined differential charges 77
- 8 Methods of increasing patient participation in payment for health care: non-provider-determined differential charges 99

Tables

- 1 Selected components of total health expenditure as percentages of
Gross National or Provincial Product 6
- 2 Impact of direct charge schemes in terms of policy objectives 112

Preface

It is difficult to pinpoint the exact origin of our interest in this topic. Not surprisingly it grew out of a more general concern with the financing of health care and the factors influencing the demand for health services. A 1976 presentation (GLS) at the Pacific Northwest Regional Economic Conference and numerous lectures at McMaster University (GLS), York University (MLB), the University of Toronto (MLB, RGE), the University of British Columbia (RGE), and the Banff School of Management Studies (RGE) served to focus our interest on the potential uses and abuses of increased patient participation in health care financing. But it was the continual recurrence of the theme in the public domain, and the prodding and questioning of colleagues on each occasion, which finally forced us to consider in detail the reasons for the topic's resilience. The final stimulus for a systematic analysis of the justifications for and effects of various proposals to increase the portion of health care costs paid for directly by patients was provided by the Taylor Report released in December 1977. This report, officially titled *Report of the Joint Advisory Committee of the Government of Ontario and the Ontario Medical Association on Methods to Control Health Care Costs*, reiterated a familiar prescription in advocating the limited reintroduction of direct charges to patients. Yet, with the exception of a brief Saskatchewan experiment with user fees, it was a prescription which had been almost universally rejected in Canada. Since the grounds for neither the prescription itself nor the previous rejections of it had been subjected to systematic analysis — theoretical or empirical — the present study appeared particularly relevant.

While the title may appear to some readers unnecessarily flippant, in fact it has very specific, intentional connotations. It is not coincidental that the re-introduction of direct charges in Ontario would benefit a number of interest groups. In fact, and this is the snare, the only group for which we are unable to discern any significant advantage is the general public.

Our analysis indicates that direct charges will benefit providers, private insurance companies, and the provincial government. Direct charges in most forms (particularly the one's most frequently recommended) will serve as an injection of additional funds into the sector and thus as a source of increases in provider incomes. Furthermore, exposure to any significant direct charge is likely to lead consumers to seek supplementary private insurance coverage. Finally, direct charges provide a means of keeping the lid on health care expenditures in government budgets while allowing total (public plus private) expenditures to rise. The snare is not likely to end up empty. It will be filled by those unfortunate enough to become ill, for it is they, and only they, who will feel the effects of direct charges.

The delusion is embodied in the recurring argument that one way to solve the 'health care cost crisis' is to make patients more aware of the cost of providing their care. The argument has two prongs. First, it is asserted that a significant amount of patient-initiated utilization consists of 'unnecessary' care which could be deterred by direct charges. Second, it is often suggested that individuals should assume greater personal responsibility for their health status through preventive lifestyle changes and should rely less on the medical profession, a transition which would be helped along by direct charges.

Upon further examination, however, we find the argument rests on very shaky ground. There is little, if any, evidence to suggest that patients are the primary generators of marginally needed care and (perhaps because of that) no evidence whatsoever to suggest that prices tend to deter that segment of care first. Therefore, while there are a number of potential avenues for introducing personal accountability, analyses of them converge upon the same conclusions — consumption of necessary care may be deterred, aggregate health care expenditures are influenced marginally, if at all, and there is little reason to believe that direct charges for health services encourage preventive self-maintenance. Moreover, deterrence of care-seeking by an individual does not in itself reduce health care use, since additional provider-generated utilization can easily offset this reduction.

The case against most forms of direct charges identified in this paper is surprisingly strong; surprising not because we derived unexpected results but in view of the resilience of proposals for such charges in Canada and the popularity of the charges themselves in the United States. Brief reflection on

the above comments, however, suggests a reasonable explanation for the persistence of direct charges in the United States, even under most proposals for national health insurance. Those groups who were identified as potential beneficiaries of any new direct charges in Ontario comprise some extremely powerful political forces in the United States. The interest of providers in a universal major-risk medical plan, for example, is understandable (or should be once this paper is read), particularly when such a plan is compared to the Canadian health insurance model. We have been significantly more successful in containing costs than our neighbours to the south, largely because of public regulation of medical fees and hospital budgets. While Canadian providers have not fared as well in fee and budget bargaining with governments, providers in the United States have had considerably less trouble raising costs. This is because a multiplicity of private insurance companies have little bargaining power vis-a-vis the providers, but can 'pass through' increased costs in higher premiums. These in turn are frequently employer-paid, in whole or in part, and their ultimate incidence is in general unknown. A national health insurance scheme patterned after the Canadian plan would close the door on a large market for those private insurers.

Last, but certainly not least, the influence of economists in helping to keep the direct charge concept alive and well in the United States should not be underestimated. It is well known that gentle massaging of relatively straightforward economic models can yield predictions of an inverse impact of direct charges on total health care expenditures. Furthermore, with the assumption that patients are fully informed consumers, those same models will also tell us that the least efficacious services will be those first deterred. But demand theory applied in this manner is devoid of any consideration of need, agency relationships, and the reasons for their importance in this market. While convenient and conventional, it may also be misleading. Providers hold the key to the dynamics of the health care market, and economic efficiency in this market is not necessarily synonymous with either consumer or provider evaluations. Technical efficiency may not produce what the public desires, and consumer preferences may not lead to optimal output levels. The persistence with which economists return to the fully-informed sovereign consumer model in studying health care may have psychological roots; it may also reflect the fine old economic concept of comparative advantage. Economists are much more adept at manipulating this model than are other health care analysts; conversely many of them are much less well informed about the institutional structure of health care delivery. It shows.

While our analysis is critical of the direct charge concept and its supporters, we do identify some limited and very selective potential roles for direct

charges. In each case the charges must be for a well-defined product about which consumers have sufficient knowledge to make informed choices. Where these conditions are met, direct charges have the potential to be a constructive mechanism for improving the efficiency of resource allocation. Even in these cases, however, the usefulness of direct charges can be compromised by the absence of other necessary market conditions or the responses of parties other than consumers.

There is, then, no support for any general presumption that direct charges are a 'good thing'. In fact, the deck seems to be stacked against their successful deployment as a cost control mechanism or as a policy intended to help achieve other reasonable social program objectives. The analysis and evidence in the following pages makes that clear — at least to us — beyond any reasonable doubt. It appears that the direct charge concept is indeed an idea whose time has gone.

Acknowledgments

This project was supported, both financially and spiritually, by the commitment of the Ontario Economic Council to furthering research on current health care delivery and financing issues. In particular, the interest expressed by Grant L. Reuber, past chairman of the Council, facilitated the timely completion of this study. Additional financial support from Health and Welfare Canada, in the form of a Research Scholar Award to Greg Stoddart and a Visiting National Health Scientist appointment at the University of Toronto for Robert Evans, is gratefully acknowledged.

Throughout the course of the work we benefited from the interest shown in this topic by a number of people with diverse backgrounds. While we shall resist the temptation to list everyone, a number of individuals and groups do deserve particular mention. Suggestions, comments, and leads on elusive literature were provided at various points by Dr David L. Sackett and Michael Mendelson, while the overall analysis benefited from suggestions by participants in both a Continuing Education Seminar of the McMaster University Department of Clinical Epidemiology and Biostatistics and an Ontario Economic Council Review Seminar. We also wish to acknowledge Dr Ed Moran of the Ontario Medical Association, whose energetic advocacy and defence of the extra-billing approach forced us to examine in detail the numerous variants of this alternative. He was particularly helpful in developing an extensive array of different justifications. Alas, all appeared on close analysis to benefit his constituents rather than the general public. The cooperation of William Mennie and Arthur Smith of the Health Economics and Statistics Division of Health and Welfare Canada allowed us to shorten the lag

normally associated with the compilation and release of Canadian health care expenditure data. We were particularly fortunate to receive a number of improving suggestions from two anonymous readers of an earlier draft of the manuscript.

Finally, special thanks go to Nancy Bishop, Mary Seminsky, Kathy McEwen, and Sharon Ineson for applying their secretarial skills to various versions of this paper, to Jean Jackson and Clive Llewellyn, who were harnessed into so many proof-readings that they could probably now, recite any page from memory, and of course to our wives, whose continued patience is an ongoing source of support and amazement.

The influence of all of the above individuals on this paper has been consistently positive. Any remaining errors or omissions are ours. In a truly co-authored work such as this, each author if given the option might very well refuse to accept responsibility for any weaknesses in the final product! But collectively we must.

1

Introduction: background and objectives

It ain't so much the things we don't know that get us in trouble. It's the things we know that ain't so.

Artemus Ward¹

The 'health care cost crisis' is one of the most hallowed and durable issues in public policy. Since the publication of the *Task Force Reports on the Cost of Health Services in Canada* (Canada, 1970) and the release of the Seventh Annual Review of the Economic Council of Canada (1970), the 'crisis' has drifted in and out of public consciousness. The Economic Council of Canada suggested that, projecting current trends, the whole of the national product would in due course be devoted to education and health services. Successive federal and provincial governments have struggled with this 'crisis' with an energy and commitment varying with the political and economic cycles. A series of typical public and quasi-public reports recommending various cost controlling measures has been collected and synopsized for the Ontario Select Committee on Health Care Financing and Costs (French, 1978).

1 Although this quotation is attributed to Ward by Darrell Huff in *How to Lie with Statistics*, neither *Bartlett's Familiar Quotations* nor the *Penguin Dictionary of Quotations* lists it. They do, however, list Henry Wheeler Shaw (better known as Josh Billings) as having said, 'It is better to know nothing than to know what ain't so' – all of which, we suppose, aptly illustrates the point!

2 Controlling health care costs

That Committee was itself the result of the latest round of 'cost crisis' discussion in Ontario. The Report of the Taylor Committee (Ontario, 1977b) made a variety of recommendations for changes in the financing and delivery of health care in Ontario. Subsequently the provincial government introduced a budget proposal for a large ($37\frac{1}{2}$ per cent) increase in the personal premiums charged for the public hospital and medical insurance programs so that premiums would cover about one-third of the cost of those programs. Public and political reaction caused this increase to be scaled down (to $18\frac{3}{4}$ per cent) and led to the creation in the spring of 1978 of a Select Committee to study the issue. At about the same time physician fee negotiations between the Ontario Health Insurance Plan (OHIP) and the Ontario Medical Association (OMA) boiled over into public debate, with the OMA arguing for an 'appropriate and reasonable' 36 per cent increase in fees and OHIP offering 6 per cent. The OMA then recommended that its members 'opt out' of OHIP and charge their patients directly according to the OMA fee schedule or one of their own choosing, leaving the patient to recoup whatever OHIP was willing to reimburse. Individual hospitals and the Ontario Hospital Association (OHA) have similarly entered the public debate, although more quietly, suggesting that current levels of government funding are not adequate to maintain the present quantity and/or quality of hospital services in Ontario. It has been suggested that individual patients should be 'permitted' to pay for additional quantity or quality of service, i.e. that hospitals be permitted to supplement their public budgets by direct charges to patients.

An old theme thus runs through most of these recent versions of the cost crisis debate — the appropriate role of direct charges to patients in the funding of health care services. The Taylor Report recommends such charges; the OMA opting-out recommendation and the suggestion of limited direct charges to patients for hospital care have the same effect. An earlier recommendation by the Ontario Economic Council (1976) for relating income tax rebates to health care use would also have the same effect, associating out-of-pocket costs to patients with their utilization of health services. The debate over increases in personal premiums was couched in much the same terms — the effects and desirability of direct contributions by users of health services.

The present study brings analysis and empirical evidence to bear on the question of the relation between health care expenditures and direct charges to patients. The issue is far from simple. The Taylor Report, the Ontario Economic Council's proposal, and the provincial government's recommendation of premium increases all argue that direct charges will tend to *reduce* overall health care expenditures, or at least moderate their growth. The OMA position, to the contrary, is that direct billing by physicians would not reduce

utilization of health services. Like the OHA, the OMA sees charges to patients as a way of *increasing* total expenditures while shifting a portion of them from public to private budgets.² Both positions cannot be correct!

This study takes the view that total expenditures for health or hospital and medical care, not simply public expenditures, are the relevant focus of policy interest. These expenditures may be financed through direct taxes, public or private insurance premiums, or direct charges to patients, but ultimately it is the citizens of Ontario who must bear the costs of health care services through whatever institutional channels are established. To an employer or an employee group, there is no obvious difference between paying premiums to a public agency or a private insurance company; the important issues are how high the premium is and what coverage it buys. Public system premiums are supplemented by direct taxes; private insurance (as in the United States) usually requires patients to make out-of-pocket payments as well. The *distribution* of costs among citizens will clearly differ depending on the balance of public tax finance and private out-of-pocket charges; but from the broader all-provincial perspective the total expenditure is what matters. Arguments for 'privatization' and 'control of the size of the government sector', which recommend shifting part of the burden of health costs from public budgets to private individuals as patients or payers of private insurance premiums, need careful scrutiny to ensure that they are not in fact proposals which increase total health costs while reducing the government's share, as the OMA opting-out proposal appears to be. Such a change would meet the economic objectives of physicians — higher fees and incomes — and provide a cosmetic reduction in government spending, but the citizens of Ontario would find themselves paying higher costs.

Starting from this global perspective, we wish to explore the various ways in which direct charges to health service users, sometimes labelled 'patient participation in financing of health care'³ can be structured, and how each

2 This position was elaborated and defended during testimony by Dr W.J. Vail (president) and Dr E.J. Moran (general secretary) of the OMA before the Select Committee on 14 Aug. 1978. (Ontario, 1978). The recurring prescription in this testimony is an injection of additional private funds into Ontario's health system through the introduction of a system of user charges.

3 In the United States, and to a lesser extent in Britain, the term 'cost-sharing' is employed instead of 'patient participation in payment for health care'; however, this may be confusing to Canadians, who normally employ the term 'cost-sharing' in reference to federal-provincial arrangements for health programs. Although revenue-raising mechanisms such as health insurance premiums have fiscal impacts on insurers and patients, the literature on patient participation consists of proposals for direct

4 Controlling health care costs

might influence total health expenditures as well as other objectives of the public hospital and medical insurance programs in Ontario. Before doing so, however, it is necessary to clear away a fairly widespread misconception concerning the health care 'cost crisis' itself.

Public perceptions do not always accord with the facts. The perception of 'spiralling costs' (whatever that means) cannot or should not be based on absolute cost levels. General inflation raises costs of everything. Rapid population growth likewise increases health care costs as well as generating the human resources to support them. The most relevant information, then, is not the behaviour of dollar costs over time, still less the share of those dollars paid through public budgets. Rather it is the share of national or provincial income devoted to health expenditure and thus not available for other things.

It is most interesting that concern over the 'cost crisis' is out of phase with this ratio. When the ratio was rising rapidly, from about 4 per cent of Canadian Gross National Product (GNP) spent on health care in the early 1950s, to a bit over 5 per cent in the early 1960s, to a peak of 7.4 per cent in 1971, little public concern was expressed. But in the 1970s, when the ratio has been more or less stable, the 'cost crisis' has been a continuing public issue. The next sections report briefly some statistics on health costs for Ontario and Canada and suggest why a policy 'crisis' may be perceived despite the stability (at least relative to previous decades) of the GNP share.

HISTORICAL PERSPECTIVE: COSTS AND UTILIZATION

Recent public perception of a problem with escalating health sector costs⁴ has been based largely on two observations: first, steadily rising health care

charges, for reasons outlined in this chapter. Direct charges are those charges to patients which are functionally related to use. They include both charges to the user at point and time of service, and subsequent charges specifically related to use. Terms commonly seen in discussions of patient participation include 'deductibles', 'deterrent fees', 'utilization fees', 'coinsurance', 'user charges' and 'copayment'. While these terms are sometimes used synonymously with both 'direct charges' and 'patient participation', they in fact refer to very specific types of direct charges which are delineated and analysed beginning in chapter 3.

- 4 There are, of course, narrow and broad definitions of the health sector. The latter has been argued to include housing and sanitation, for example. We adopt a much narrower definition in which we deal primarily with hospital and medical costs. Other related services such as dental and pharmaceutical care will at times be introduced for illustrative or data base reasons. Here the reference is to total health expenditures (Canada, 1975), the trend in which is dominated by hospital and medical services which constitute approximately two-thirds of the total.

costs throughout the 1960s, as shown in Table 1, reaching a peak of 7.4 per cent of GNP in 1971 and, second, an apparent resumption of the escalation after a 1972-4 pause. Coming after the steady cost increase of the previous two decades, this pause appeared to lend some support to those who argued that the previous expansion had been a once-for-all accompaniment of the extension of private, then the introduction of public, hospital and medical insurance. The alternative view, that continued cost escalation was inherent in the structure of the health care delivery system, seemed to be refuted. Despite a continuing series of studies diagnosing and documenting the essentially unmanaged nature of health care delivery and recommending more or less sweeping changes to improve efficiency and effectiveness, no significant changes were made on the supply side of the hospital and medical industry to correspond to the almost total displacement of the market by public payment mechanisms on the demand side. Yet for a couple of years after the last province joined the medical insurance system in 1971 it appeared that, despite the failure of federal and provincial governments to address the hard questions of restructuring the delivery system, the problem of steady escalation in the health care share of the Canadian economy had disappeared. The Economic Council of Canada (1970) to the contrary, health care would not absorb the GNP.

A number of interesting trends are evident in Table 1. First, the pattern of hospital plus medical expenditures as a percentage of Ontario Gross Provincial Product mirrors closely the comparable national pattern. This is not particularly surprising when one recalls Ontario's approximately 38 per cent share of national expenditure over this period for those service categories and the similarity of all the provincial plans. Second, the national hospital plus medical ratio series closely parallels that for total health expenditure, again not surprisingly in light of the dominant role of those two components in the construction of the total expenditure series. Perhaps more interesting is a third comparison, between Canada and the United States. The dramatic post-1971 levelling off of health expenditure as a percentage of GNP in Canada has not happened in the United States. Whereas the most recent figures put Canada's ratio at the level first reached in 1970, the American ratio continues to set new record levels, thus widening the differential. As will be noted again later, this makes any argument that universal health insurance itself is the causal factor underlying health care cost escalation difficult to sustain.

While the limited figures available for years immediately after the pause (post-1974) suggested a possible resumption of cost escalation, it is now clear that health care costs have consumed a roughly constant share of national

6 Controlling health care costs

TABLE 1

Selected components of total health expenditure as percentages of Gross National or Provincial Product

	Ontario hospital plus medical expenditure as a percentage of Gross Provincial Product	Hospital plus medical expenditure for Canada as a percentage of GNP	Total health expenditure for Canada as a percentage of GNP	Total health expenditure for the United States as a percentage of GNP*
1960	3.1	3.1	5.6	5.3
1961	3.2	3.3	5.9	5.5
1962	3.3	3.3	5.9	5.6
1963	3.4	3.5	6.0	5.7
1964	3.4	3.5	6.0	5.9
1965	3.4	3.6	6.1	5.9
1966	3.4	3.6	6.1	5.8
1967	3.7	3.9	6.4	6.2
1968	3.9	4.1	6.6	6.5
1969	3.9	4.2	6.8	6.7
1970	4.2	4.4	7.1	7.2
1971	4.4	4.6	7.4	7.6
1972	4.3	4.5	7.2	7.8
1973	4.0	4.3	6.8	7.7
1974	4.0	4.2	6.7	7.8
1975	4.3	4.5	7.1	8.4
1976	4.2	4.4	7.1	8.6
1977				8.9

*fiscal year basis

SOURCE: (Canada 1973a, 1973b, 1975, 1978) as well as forthcoming figures for 1974-6 and revised figures for earlier years from Health Economics and Statistics Division, Health and Welfare Canada; Mueller and Gibson (1976), United States (1977), Ontario (1975, 1977a), as well as provincial GPP estimates for 1975-6 from Ontario Ministry of Treasury, Economics, and Intergovernmental Affairs.

resources since universal hospital and medical insurance became complete in 1971. It may be seen as somewhat puzzling, therefore, that the widespread belief in the existence of a health care cost crisis remains.

Closer examination suggests at least three factors which may be responsible for keeping the subject of health care costs in the public eye. Foremost among these has been the increased concern of provincial governments with the level of expenditures for insured services in the period leading up to and

following implementation of the Established Programs Financing Act in April 1977. With the demise of the fifty-cent dollar, with federal transfers for health care no longer being tied directly to levels of provincial spending, provincial authorities are paying increasing attention to matters of efficiency and areas of potential cost savings. Moreover, the slowing of growth in the economy generally and the increasing public concern over the size and scope of governments has led to an increase in the political 'cost' of a dollar of tax revenue. Increased intensity of competition within government for such funds and Treasury pressure to contain spending lead to close scrutiny of very large expenditure sectors such as health care. To some extent, then, generating public consciousness of a health care 'cost crisis' may be a necessary tactic to maintain a political climate in which restraint on health expenditure can be achieved. Because the crisis is perceived, it can be averted; if people did not think there were a crisis, there soon would be one.

Partly as a result of this scrutiny and partly because of the erosion of physicians' unusually high post-medicare real incomes, resulting from negotiated fee increases having fallen short of the increase in the general price levels since 1971, professional associations have increasingly argued their case for higher fees in the media and attempted to gain popular support by direct communication with patients.⁵ Insofar as the relative cost stability of the 1970s has been achieved by holding physicians' incomes and hospital and other institutional growth below the rates which had come to be taken for granted during the previous two decades, there has been a steady build-up of political pressure from both professionals and hospitals to have these restraints relaxed. If this pressure is successful, the 'cost crisis' would re-emerge in public or in private budgets. Whether this occurs depends on whether the providers of health care revise their expectations of the resource entitlements of the health system and on the strength of political resistance.

- 5 For example, early in February 1978 the Hamilton Academy of Medicine supported a campaign in which some local family physicians sent letters to approximately fifty thousand families in an attempt to build public support for higher fees (*Hamilton Spectator*, 1 Feb. 1978). This followed closely upon an announcement by the Ontario Medical Association that it had approved a 'realistic' fee schedule, calling for a 36 per cent overall increase in fees. (The Ontario Health Insurance Plan subsequently approved an average fee increase of 6½ per cent, thereby abandoning the correspondence between the OMA fee and OHIP benefit schedules.) Placing a 'realistic' value on a service is equivalent to establishing an appropriate relative real income for the provider. As the figures below demonstrate, current physicians' incomes have recently fallen from their immediate post-medicare peaks but are still relatively high if viewed in a longer historical perspective:

8 Controlling health care costs

Finally, numerous recent inquiries by concerned American health professionals, administrators, and observers attempting to glean lessons from experience with universal public health insurance have guaranteed continued discussion among their Canadian counterparts regarding deficiencies in and disappointments with several aspects of the Canadian system.⁶

In addition, however, there appears to be a growing dissatisfaction with the level of health in relation to the level of expenditure on health care. A consensus is emerging among analysts and administrators of the health care

Ratio of physicians' net annual receipts to annualized industrial composite average weekly wage (selected years)	
Year	Ratio
1957	3.78
1960	4.15
1963	4.37
1966	4.83
1967	5.08
1970	5.42
1971	5.70
1972	5.36
1973	5.14
1974	4.75
1975	4.46

Source: Figures for 1957-71 are from Evans (1975); subsequent years are from Evans and Wolfson (1978).

While more recent data are unavailable, the combination of limitations on numbers of immigrant physicians and higher rates of increase of negotiated fee schedules have probably arrested this downward trend.

- 6 Perhaps the most comprehensive inquiry was the 1974 Symposium of the Sun Valley Forum on National Health, which resulted in an edited publication of the symposium proceedings (Andreopoulos, 1975). More recently, numerous articles (e.g. Marmor and Tenner, 1977) attempting to distil lessons from the Canadian experience have appeared in American journals. Several of the recent visits by Americans involved in the debate over national health insurance are summarized by Korcok (1978). In fact, the raging public debate over a national health insurance program for the United States ensures even more public exposure to cost 'information' there than here, and may be responsible for a fourth factor. Canadian readership of American newspapers and magazines and viewing of American television ensures ample Canadian exposure to the trends in the United States. The fact that costs there have not exhibited the same recent stability as the share of GNP in Canada (as illustrated in Table 1), coupled with the many parallel features of the two economies, could induce erroneous extensions to the Canadian experience by those unfamiliar with Canadian figures.

system that we should, and could, be buying higher levels of population health status with current resource commitments or that existing levels of health status could be achieved and maintained with lower resource expenditures. While part of this discrepancy between expected and actual health status may be attributed to the limited scope for the health care delivery system in combating environmental and lifestyle-related conditions (Lalonde, 1974; McKeown, 1976), a significant part may also be a result of inappropriate consumer demands, ineffective services, and technical inefficiency in the production of medical and hospital care. To the extent that the latter claims are substantiated, proposals to control health care costs must operate not only on levels but also on specific patterns of utilization and expenditures. But whether increased consumer financial involvement is the best or even an appropriate vehicle for achieving this result has yet to be demonstrated.⁷

- 7 There is a quick retort to the charge of health system inefficiency expressed in this paragraph: health care has (little or) nothing to do with health. Health depends on all the various factors discussed in Lalonde (1974), and the role of health care is minimal. Health *care* really serves to comfort the afflicted, to relieve anxiety, and generally to sustain patients and improve the quality of their lives as they cope with the inevitables of the human condition – disease and death. Thus it is quite inappropriate to judge the efficacy of the health care system or gauge appropriate levels of health care by reference to the health status of the population.

This position strikes us as inconsistent with the historical stance of health care providers. In negotiating with society in general (either individual patients, communities, or governments) for a larger share of social resources to be devoted to health care, providers have always marketed their services in terms of health status. ‘Give us more resources and we will cure . . .’, or, alternatively, ‘If adequate resources are not forthcoming, the hospital or the medical care system will not be able to assure adequate care.’ Quite obviously, the threat (however loosely specified) of illness, incapacity, and death is a potent negotiating tactic – physician and hospital strikes are not feared by the community merely because of the inconvenience. It thus seems a little unfair for health care providers *now* to respond that health care has nothing to do with health and therefore that their efficiency and effectiveness should not be judged against that standard. (What is the appropriate standard then? If it is consumer willingness to use physicians’ services, the justification for professionalism is lost.)

Unfair or not, the position deserves further scrutiny. One cannot reject it just because it is inconsistent with earlier marketing claims. At a more substantial level, however, the proposition seems faulty because it takes too narrow a view of health status. By an inefficacious procedure, we do not mean merely one which fails to prolong life. Rather we mean one which, for a particular patient, has been found to have no positive effect whatsoever. Surgery, for example, which makes the patient more comfortable, although not prolonging life, is not unnecessary; that label is

10 Controlling health care costs

HISTORICAL PERSPECTIVE: CONSUMER PARTICIPATION

A significant part of the motivation for public health care insurance in Canada was a perceived inequity in access to private insurance. While private insurance fulfilled the role of shifting risks, it did so selectively, and often those persons most at risk and least able to pay were the ones who found it most difficult to enrol in a private plan. Thus, one of the intended and principal roles of a public plan was to assure equal access to services, regardless of health status and/or income level.

Even as the debate over national health insurance was being initiated, recognition of the potential cost implications of such a scheme in various guises was leading to consideration of ways to involve patients in financing. Equal access implied no consumer charges (though the converse does not follow), which in turn suggested potential cost control problems – problems which would land squarely on the shoulders of the governments involved. The

reserved for procedures which turn out not to make the patient feel better either. Of course, we have to abstract from the placebo effect; moreover, one can hardly count as a benefit the relief of patient anxiety which was created by the physician's statement that surgery was needed. That would create a totally circular concept of 'necessity'.

A proper definition of health status *must* include quality-of-life factors as well as the crude morbidity/mortality measures, and it is in relation to this broader definition that a number of interventions have been judged inefficacious. It must be remembered that any medical intervention by itself is likely to have a negative impact (in the short run at least) on quality of life. This, as well as the resource cost of intervention, must be balanced by some subsequent objectively identifiable positive gains in life quality, or else the procedure is inefficacious.

Moreover, the more broadly based our concept of health status, the more serious become the issues of technical inefficiency in health care delivery. The problem of wasted resources results not merely from doing things which should not have been done (because they did more harm than good) but also from doing useful things at unnecessarily high cost. The more one justifies health care in terms of its 'carative', anxiety-reducing, quality-of-life contributions rather than its curative, defence-against-death roles, the more questionable is the emphasis on technologically intensive and highly trained service. Serving as a sympathetic listener does not require seven to ten years' training and a sophisticated technological support system.

It thus seems that those who argue that health care has nothing to do with health in defence of the *existing* delivery system are trying to have it both ways. As a curative defence against disease and death, the present system must be judged in terms of its efficacy in influencing health status. As a carative system, it must be judged in terms of the efficiency of the pattern of resources deployed for this purpose. Either way, there is evidence of considerable room for improvement.

issue of consumer participation through such schemes as major-risk insurance dates back, then, to pre-Royal Commission (Canada, 1964) debates over the design of a public program (for example, see Anderson, 1952, with comments by Willard and Taylor).

Before the release of the Royal Commission Report there was some degree of unanimity amongst medical associations and private insurers as to the position the Report would adopt (LeClair, 1975). In particular, it was felt that some patchwork system integrating public and private insurance plans would be recommended, with a direct charge structure built into the private plans. In that way, those excluded from private plans as bad risks would be covered by a tax-financed plan, while the great majority of consumers would be responsible for some combination of premiums and direct charges. In fact the establishment of the Royal Commission was in part a response to pressure from the private insurance industry, whose members believed that a thorough and complete investigation would reject 'state hospitalization' along the lines of provincial plans then in existence (as D.E. Kilgour stated to the annual meeting of the Great West Life Assurance Company, reported in the *Vancouver Province*, 7 Feb. 1956, 14). Quite the opposite happened. After extensive discussion and deliberation, the Commission rejected consumer participation in payment and offered instead a proposal for first-dollar universal insurance. Direct charges were conspicuous by their absence. Evidently equal access was a higher priority at that time than potential cost problems, and direct charges to consumers were not viewed as a particularly effective way of addressing those problems.

The fact that direct charges were considered unattractive by the Royal Commission has not prevented limited subsequent implementation. These initiatives have only provided support for those opposed to direct charges. For example, Saskatchewan's flirtation with deterrent fees for physicians' services from 1968 to 1972 did indeed result in lower overall utilization (Beck and Horne, 1978). The distribution of that decline, however, was in direct conflict with the equal-access objectives of the universal plan. In particular, the decline was concentrated within low-income groups (whose members tend also to be relatively elderly). In general, the introduction of direct charges has either induced no decrease in utilization and expenditures or has had unpleasant side effects.⁸ In addition, such American studies as Roemer et al. (1975), reworked with similar results by Helms et al. (1978)

8 A comprehensive documentation of Canadian experiences with consumer direct charges is presently being compiled (Badgley, 1979).

12 Controlling health care costs

and Hall (1966, 1974), suggest that cost control through patient participation may be little more than an illusion.

The persistence of the concept of direct charges despite the theoretical and empirical opposition to it is somewhat puzzling. The traditional deterrent charge has been joined in recent years by other policy options directed at the consumer. The example of extra-billing by opted-out physicians, noted earlier, is of particular interest to Ontario residents. The current American debate over a national insurance plan in which coinsurance, deductibles, and major-risk insurance are commonly recommended options, recalls the Canadian discussions of the 1950s. Recent changes to the pharmaceutical dispensing function in Saskatchewan (about which more will be said in chapter 4) are an attempt to involve consumers on a limited basis in expanding the market share of efficient providers. Finally, in addition to extra-billing by physicians, the possibility of reintroducing some form of deterrent charge to patients in Ontario for hospital and/or medical care has recently received attention in some quarters.⁹

Why is requiring patients to pay directly for care such a persistent recommendation when an apparent cost crisis looms? One reason is the argument that direct charges to consumers foster awareness about the costs incurred on their behalf (the visible-link argument). It is not clear, however, what impact awareness has on cost-generating behaviour. A more likely reason is the standard supply-demand market model of any first-year economics course, which suggests that imposing prices on users of health services automatically reduces quantity demanded (and thus utilization) and thereby limits costs. (It is after all a fairly elementary proposition that health care costs are the product of the volume of services produced and the average price/cost of these services; control of total costs must then result from control of one or both of these components.) Any concern about denying 'needed' services to those who cannot pay can be alleviated simply by accepting as correct the claims of physicians' organizations that their members are overburdened with trivial and frivolous complaints.

On a more sophisticated plane are the data presented in the federal government's White Paper (Lalonde, 1974), suggesting that additional health care may contribute little to our health status, and the extensive and growing medical literature on costly and inefficacious procedures to which Cochrane (1972), Illich (1975), McKeown (1976), and a forthcoming federal report of the Task Force on the Periodic Health Examination provide extensive re-

9 An illustrative example is the Taylor Report (Ontario, 1977b).

ferences. Furthermore, the stress on environmental and lifestyle factors as influences on health status, evident in the White Paper and in Morgan (1977), can readily be blurred with notions of a need for greater 'personal responsibility' in health care, notions more popular in the 1970s than a decade ago. From this position it is not too long a step to the view that if people are ill it is their own fault, and they ought to pay for it. A set of problems and concepts thus seems to support the idea that some form of direct financial responsibility by patients for the cost of their health care would improve the efficiency and lower (or at least moderate the growth of) the costs of the present system.

But the 'blurring' necessary in the latter chain of arguments and the implicit assumptions upon which the 'pop economics' model is necessarily based are difficult to reconcile with the complexities which characterize the health care market. Far more careful consideration of the role of those characteristics must be incorporated if one is to be able to undertake a realistic assessment of consumer-participation policy options.

ON CURRENT SNARES AND DELUSIONS

The present paper explores the benefits of a general policy of consumer participation in health care financing. It focuses exclusively on *direct* charges to patients. Premiums are thus not considered. Why not?

As noted earlier, the revenues raised through premium collection in Ontario are not specifically earmarked for reimbursing health care suppliers. Compulsory premiums are a poll tax, and like any tax are intended primarily as a fund-raising mechanism.¹⁰ They are not direct charges and are not in any way linked either to one's actual record of utilization (experience-rated) or to one's expectation of future use. (Since premiums are set well below the average expected cost of care and supplemented by taxes, and since alternative private insurance is not permitted, failure to enrol in OHIP cannot be economically rational even for those not compelled to join.) The deterrence effect of any scheme in which patient payment is not a function of use is not

10 It could be argued that taxation is also intended to redistribute income. But redistribution is a separate objective which can be introduced on the tax or expenditure side; in effect the distinction is between tax incidence and net fiscal incidence.

14 Controlling health care costs

obvious and not likely to be significant.¹¹ We risk little, then, in dismissing premiums as a reasonable cost-containment mechanism involving the consumer. With respect to arguments that consumer participation facilitates consumer awareness, there is no evidence to suggest that premiums are better able to inform the consumer about the cost of rendered services than any other form of tax (income, sales, payroll) which might be labelled as a source of health care funding. Revising premiums does nothing to change the allocation of resources to or within the health sector; it merely redistributes the financial burden of a given resource allocation pattern. A necessary though not sufficient condition for consumer participation to have an impact on resource allocation is that the scheme in question involve direct charges.

The case for the reintroduction of direct charges rests, as suggested above, on a grossly oversimplified structure of implicit assumptions about the way the 'medical marketplace' functions. While it is true that total health care costs are the product of volume of services utilized and average cost per service,¹² patient behaviour (however affected by direct charges) is only one component influencing utilization. One cannot realistically expect to affect one element without inducing countervailing shifts in the other components and/or in unit costs. At least as important as patient behaviour is the behaviour of the provider of services, and this behaviour is likely to shift utilization in a direction opposite to any independent shift in patient behaviour. The result of such interconnected shifts in behaviour can quite easily be an *increase* in total health care costs — a possibility which has more than theoretical interest when we observe (Table 1) that health costs are substantially higher in the United States, where patient participation in hospital and medical funding is widespread, than in Canada, where such charges are much more limited. Moreover, again as shown in Table 1, relative cost escalation was substantially more rapid in Canada in the pre-medicare period of mixed public and private insurance and direct charges than it has been since 1971. The market for health care is very complex and interdependent, and any analysis which fails to take into account the interests and

11 In fact, premiums or other labelled taxes may have quite the opposite effect if the 'we've paid for it, why not use it' attitude predominates. There may, however, be a 'visible link' between tax changes and utilization behaviour at the collective rather than individual level. Although no individual can expect any change in premiums as a result of reduced personal utilization, the residents of a district or municipality may come to see a connection between the protection of surplus hospital facilities, for example, and subsequent tax increases. Or they may not.

12 'Services' is here meant to include all health care goods and services.

behaviour of the physician (or, where relevant, the dentist) is almost certainly erroneous. It would be *Hamlet* without the prince of Denmark.

Failure to recognize the nature of the utilization process and the role of provider influence is a common pitfall in discussions of direct charges. Compounding the problem is the missing link in this market between increased production of health care and increases in community health status. The sales revenue generated by the health care industry for its products, out of which incomes and expenses are paid, is the cost of health care to the rest of society. Increased production means increased costs, but, since more health care is not necessarily better than less from a community health status vantage point (and at the margin), increased costs need not correspond to higher levels of output. This ultimately becomes an empirical question plagued by output measurement problems, but sufficient evidence does exist to suggest that much of the industry's production is inefficacious, unevaluated, or unnecessarily costly for the problems addressed.¹³

To investigate the potential benefits of consumer direct charges and to attempt to distinguish plausible arguments from unfounded claims, we shall survey and analyse the range of different types of charges to patients which have been used or suggested for use in the Canadian health care system. Each such charge has, like any other price charged in a market system, a dual effect. It transfers wealth from the person paying the charge to the person or agency receiving it. But it also provides a signal and an incentive to the payer to modify his behaviour in some respect. The incentive may or may not elicit a response, but if it does that response will have further wealth transfer and incentive effects on other participants in the health care market, and they too will respond in turn. Any recommendation for or against direct charges must thus be built upon an implicit or explicit structure of assumptions about what these various behavioural responses will be. By making this structure explicit, one can assess its inherent plausibility and empirical support. In many cases, for instance, 'pop economics' arguments rest on a detailed structure of assumptions which have become implicit (if not forgotten) through long use and habit. Their relevance to a particular policy situation is not automatic and should not be taken for granted.

The development of an alternative framework within which to assess the effects of direct charges is not straightforward. Discussions of patient partici-

13 The literature here abounds. Evidence of unnecessary surgery (Roos et al., 1977a, 1977b, 1978; Roos, 1977; Vayda, 1973; Bunker, 1970; Bunker et al., 1977), inappropriate drug therapy (Walker, 1971; Silverman and Lee, 1974; Illich, 1975), and inappropriately placed hospital patients (Gross, 1978) comprises a limited selection.

pation in financing health care have been subject to a great deal of confusion because of the wide variety of different types of schemes proposed, the diverse and often conflicting objectives they are meant to serve, and the rather extensive assumptions about the structure and behaviour of the health care industry necessary to bridge the gap between proposals and objectives.

We attempt, first, to develop a taxonomy of different types of proposals for direct charges. For each proposal within that taxonomy we specify the assumptions about the health care industry that are necessary to predict its impact. Drawing on empirical and other information, we attempt to assess the plausibility and consistency of these assumptions and thereby evaluate the probable effects of each option on a set of policy objectives, outlined in the next chapter, which are considered representative of those of a public health insurance plan.

There are many ways in which one might distinguish between types of direct charges. We have adopted the framework illustrated in Figure 1 in which distinctions are maintained in two dimensions – whether the charge varies with different providers and who determines what the charge will be. A detailed justification for this framework is provided in chapter 3; suffice it to note here that the uniform/differential pricing distinction is felt to be important because it relates directly to the scope for attaining one of the objectives identified in the next chapter, that of technical production efficiency. This form of efficiency refers to the production of a particular health service at minimum cost. It is to be distinguished from the broader question of whether that service should be provided at all. The identity of the party determining the charge is significant because the pattern of differential charges (across providers) which will emerge if providers set their own rates may be quite different (and will serve different purposes) from a charge structure directly influenced by third parties.

Proposals involving charges which are uniform across providers, those in groups A and B of Figure 1, are examined in detail in chapter 3. Differential charge proposals, those in groups C and D of Figure 1, are analysed in chapters 4 and 5.

Through analysis of the various schemes for patient participation identified in chapters 3 to 5, it is possible to identify a set of preconditions which, if satisfied, might permit certain forms of direct charges to have a significant impact on the specified policy objectives. In chapter 6 we summarize the anticipated scope for patient participation and its likely impact in view of these preconditions.

FIGURE 1

Methods of increasing patient participation in payment for health care through direct charges

		Charges determined by	
		Non-provider	Provider
Level of charges across providers	Uniform	A	B
	Differential	D	C

Note: Non-providers are third parties – governments or insuring agencies; providers are individual practitioners, institutions, or associations.

2

Health system objectives and the role of health insurance

'Would you tell me, please, which way I ought to walk from here?'
'That depends a good deal on where you want to go', said the Cat.

Lewis Carroll
Alice in Wonderland

Comprehensive analysis of the effects of consumer participation in payment requires both an examination of various participation schemes and a set of assumptions about the behaviour of patients and providers. But first the criteria or public objectives against which any scheme is to be evaluated must be identified or agreed on. Often these objectives are fuzzy, inconsistent, or unrealistic. One may admire a health care system which attempts to 'maximize health', 'minimize cost', or better still 'maximize health at minimum cost', but these terms do not describe any current system. Health maximization would likely imply an even larger share of national resources committed to this sector. Cost minimization would mean eliminating most, if not all, services. Finally, any attempt to maximize one variable while minimizing another cannot succeed as long as the two are directly and functionally linked.

Such 'public objectives' in the health sector are obviously facile characterizations. More formal representations of the social welfare function are also easy to come by and entertaining to manipulate but lacking in operational content. One could of course argue, as does Culyer (1976, 1978), that the purpose of the health care system is to maximize the health status of the

population within the budget for health care made available by the state. Health status would be measured by some index constructed by summing across the various dimensions of health and across members of the population, using socially determined weights. Such a proposal, though radical, turns out to have attractive features as well as problems, but it cannot serve as a guide to policy in the area of charges to patients. The issue of charges is bound up with the broader question of the level of resources to be devoted to health care, assumed to be predetermined in Culyer's formulation, as well as with the practical problems of how in an imperfect world to attain the level of technical efficiency in health care delivery that is also presupposed in his approach. It turns out that one *can* make an argument for certain kinds of direct charges to patients as potential incentives to promote this sort of improvement in technical efficiency, but only under carefully specified conditions and definitions of the 'product' to which the charges are linked. As we shall find, if the 'product' charged for is inappropriately defined, the case for direct charges collapses.

Defining public objectives is not easy. Nevertheless, every health care analyst undoubtedly has conveniently at hand at least one favourite specification of the proper goals of public programs, whether or not he will confess to its existence. In this paper we suggest a set of four policy objectives appropriate to public health care insurance or delivery programs. These goals have been employed before (and developed in more detail) in an analysis of dental and pharmaceutical insurance (Evans and Williamson, 1978), and they seem equally suitable to the question of charges to patients. They divide naturally into two pairs: insurance objectives and health care system objectives.

HEALTH INSURANCE OBJECTIVES

Health insurance objectives are independent of the quantity of health care utilized or the technical efficiency with which it is produced. They consist of *risk reduction* and *wealth transfer*. Consider for a moment a public program of hospital insurance that (miraculously) had no effect on hospital costs per patient day, per patient stay, or per procedure or on the level of utilization of hospital care. If such a program were financed by differential premiums exactly tailored to each individual subscriber's expected hospital expenses during the period of coverage, its benefits would be purely those of risk-reduction. The subscriber would exchange a situation of a large but uncertain expense, such that his net wealth at the end of the period of coverage is

unpredictable, for a small and certain premium. End-of-period wealth is now predictable, and if the premium is actuarially fair (equal to the size of the uncertain loss multiplied by its probability of occurrence) the average or expected net wealth will be the same with and without insurance. The costs of administering the program will naturally have to be included in the premium, so that in fact expected net wealth will always be lower in the insured case, but this cost is offset by the reduced risk or the enhanced predictability of end-of-period wealth.¹

As long as each subscriber's premium is adjusted to his own expected expenses, such a pure insurance program would involve no *ex ante* transfers of wealth between subscribers. *Ex post*, or at the end of the coverage period, those without expenses would clearly have transferred wealth through the program to those with expenses. But except for the administration costs no subscriber's expected wealth would rise or fall immediately after joining a pure insurance plan. If on the other hand the plan were financed through taxation (which includes compulsory premiums, nominal or *de facto*, unrelated to each subscriber's expected use), it would also tend to transfer wealth *ex ante* from those with low probability of use to those with high probability.² Such transfers before actual claims experience becomes known mean that the expected net wealth at the end of the coverage period is raised for some (expected high-use) subscribers and lowered for other (expected low-use) subscribers as soon as they participate in the plan. These wealth transfer effects before the actual outcome or program experience becomes known are considered an additional benefit if those with high probability of use are the old, the indigent, or those otherwise considered by society as deserving of transfer. In fact these groups are relatively very high users of hospital services. If offered a program without such wealth transfer features they would be unable to participate, because their expected use is so high in relation to their resources that they could not afford actuarially fair premiums — hence the inability of experience-rated private insurance, before

1 This benefit can be easily shown to be more significant for high-cost, low-frequency risks and less significant conversely, for such services as pharmaceutical and dental care (Evans and Williamson, 1978, appendix A).

2 One may not find decreases in the expected net wealth position of *all* individuals with low probability of use since, depending on the progressivity of the tax system, some expected low users might also pay very little into the plan. Similarly, under a progressive tax system there may be little change in the expected net wealth position for high-income/high-use individuals. In this situation, the wealth transfer would be from low users with high income to high users with low income, rather than simply from low to high users.

22 Controlling health care costs

the Canadian public programs were implemented, to extend coverage to these groups.

Both of these benefits, it should be emphasized, are analytically independent of any program impact on population health levels or expenditures, because they neither depend upon nor result from changes in utilization or unit costs. (Such changes could, of course, intensify or dilute the wealth transfer effect.) Furthermore, as the following chapters demonstrate, any policy of reintroducing direct charges would tend to increase risks, and most such policies would reverse the wealth transfer effects discussed above even if they had no impact on utilization or costs.

HEALTH CARE DELIVERY SYSTEM OBJECTIVES

Health system objectives focus on the issues of utilization and costs. Public policy in health care is not intended merely to reshuffle financial claims, though that may itself be useful. It is intended to influence *levels and patterns of utilization* of health services, with the ultimate object of improving the health status of the population. In the early and naive days of health insurance, this was expressed as lowering the barriers to care and meeting the iceberg of unmet need. More health care was expected to lead to more health and was thus a Good Thing in and of itself. As noted above, more recent attention has focused on forms of health care which are of marginal benefit or positively harmful to health, such as unnecessary surgical procedures and overused prescription drugs.³ There remains a social interest in health care utilization per se, but it is no longer simply an interest in 'more'. A delicate adjustment — more of some things, less of others — in patterns of care is by implication now the appropriate goal of social policy.

- 3 As discussed in chapter 1, note 7, benefit here refers to a relatively comprehensive definition of health which includes quality-of-life factors. In this broader sense it is clear that health care is expected by the general society to improve 'health'; that is why individuals use it and societies pay for it. Unnecessary or inefficacious procedures or drugs are so judged not because they fail to 'cure' in some absolute sense but because they cannot be shown to yield any objective improvement in the patient's quality of life. 'Objective' is emphasized, because subjective improvement may result either from placebo effects or relief of general anxiety or recovery from a perceived deterioration generated by the physician's recommendation of intervention. If the performance of inefficacious procedures or the administration of inefficacious drugs is justified on the grounds of subjective improvements of that type, the distinction between professional practitioner and quack or witch doctor literally vanishes. The policy implications are quite significant!

Since charges to patients would affect these patterns of utilization as well as aggregate utilization levels, both aspects must be included in an evaluation. As Culyer (1976) points out, medicine is now too important to leave solely to doctors. It is precisely this public interest in patterns and levels of utilization which tends to draw policy-makers into the practice of medicine as they attempt to influence what is produced.

The social interest in health care patterns and levels may be expressed *through* government institutions and interventions, but it does not arise as a *result* of government's increased intervention in health care. On the contrary, the direction of causality is exactly the reverse. Government's role in health care delivery has grown in response to our interest, as members of a common society, in the health of all of us; it reflects our interdependence. Each of us, in actively or passively supporting public intervention, expresses an interest in the health of our neighbours. This interest in health has generally been operationalized as an interest in access to health care of appropriate quality, but it is increasingly evident that general increases in utilization of health care can have negative or neutral effects on health as well as positive ones. The social interest in health is no longer effectively expressed as a desire for 'more' health care, but rather as a concern for the mix and balance of different types of services as well as for their overall quantity. This interest cannot be served by social institutions (public or private) incapable of influencing utilization mix and thereby the efficacy of this mix.

The second health system objective reinforces the tendency towards public policy involvement, because the technical 'how' of production is also a policy issue. Society clearly has an interest in promoting the maximum *technical efficiency* of the health care production and delivery processes, that is, in obtaining any particular quantity and mix of health services with minimum resource commitments. Wasted resources in the health care industry, as in any other, lower the overall welfare of society. As in the case of the social interest in mix and levels of service utilization, the interest in technical efficiency may be expressed through government intervention but is not a result of government intervention. The technical efficiency objective is distinguished from that of utilization by the notion that useless or harmful services, if produced, should be produced with maximum efficiency or the minimum use of scarce social resources. This is of course true for efficacious services as well. The technical efficiency objective may be more clearly identified if government is paying all the health care bills, but it would be just as important in a completely privatized system. Wasted resources are a loss to society, whether as unnecessarily high taxes, private insurance premiums, or privately paid medical or hospital bills. Public regulatory or competitive

private market institutions are thus alternative social mechanisms attempting with varying success to encourage the achievement of technical efficiency.

In most industries the built-in incentive structure of the private competitive marketplace can be relied upon to promote efficiency at least as well as any public policy-maker might do. But such conditions may not obtain when competition is absent or suppressed, and public intervention may be called for to induce or restore efficient behaviour.

In health care, competitive behaviour in the conventional economic sense does not seem generally feasible (it is conspicuously absent at present), and other forms of public intervention may be necessary to promote technical efficiency. At the same time, as noted one might wish to look for ways of restructuring the organization of health service delivery and redefining the 'product' so that conventional competitive forces could play a useful role. Havighurst (1974, 1977) in particular has urged this policy in the United States. More recently, Enthoven (1978) has suggested that competition between groups of providers offering to provide all 'needed' services to groups of individuals for a fixed annual capitation payment could induce both increased technical efficiency of provision and a more efficacious pattern of service utilization, a concept discussed further in chapter 4. In such a restructured environment, charges to patients might play a significant role in encouraging improved efficiency. But charges would no longer be for particular health services. Rather, they would be fixed annual payments for an unspecified service package to be determined by the provider as need arises. As Havighurst has emphasized, the necessary institutional restructuring would be significant. One cannot turn back the clock merely by reimposing patient charges within the existing health care delivery system.

In summary, then, the policy objectives against which consumer participation proposals are to be evaluated in this report are, first, the reduction of financial risk to which individuals are exposed as a result of unpredictable illness; second, the facilitation of 'appropriate' wealth transfers, that is, redistribution of the financial burden of illness more in proportion to ability to pay and less in proportion to illness incidence; third, the modification (increasing or decreasing) of the levels and types of health care which people use in order to improve health status; and, fourth, the encouragement of technical efficiency in the production and delivery of health services.

As will become apparent in the analyses that follow, even though proposals may differ in their effectiveness in light of any given objective, fundamental analytic themes and distinctions underpin all the proposals and emerge throughout the remainder of this study. The following chapter begins by considering, in some detail, these themes and distinctions.

3

Charges uniform across providers

The notion that hospital use has risen because third-party coverage has led to patient-originated overuse is as illogical as contending that prisons are overcrowded because there is no copayment for their use.

M. Blumberg (1977)

As noted in chapter 1, our attempt to evaluate any specific proposal for direct charges to patients in terms of the four policy objectives identified above requires us to separate charges which are uniform from those which vary across providers.¹ Flat charges of \$4.00 per hospital day, or \$2.50 per physician office visit are examples of the former, while charges deter-

1 Here and henceforth we use the term 'provider' to refer to an individual or institution providing to a patient a health service (or commodity) for which the patient could be charged directly. Hospitals, clinics, physicians, dentists, and pharmacists are all examples of health care providers. Hospital or clinic employees, even if physicians, are not providers; it is the institution which provides the service through its employees. Public agencies such as health units are also providers insofar as they provide specific services to identified patients. In general discussion, 'provider' will frequently refer to the physician (more correctly the medical practice) unless otherwise specified. Although there are obviously important distinctions between different classes of providers in behaviour, objectives, and constraints, providers can usually be treated as similar at our level of analysis as well as in their economic interactions with patients and public insurers.

26 Controlling health care costs

mined at the discretion of physicians who are permitted to extra-bill are among the latter. The importance of this distinction lies in the fact that uniform charges do not possess the potential of differential charges to encourage improvements in technical efficiency.

Although uniform-charge schemes will always have risk and wealth transfer effects, their primary impact on the health system will be limited to changing, presumably reducing, the quantity of health services demanded.² For any mechanism of consumer payment to influence the technical efficiency of suppliers, it must be part of a price-competitive system in which the choices of informed consumers reacting to price signals (direct charges) allow more efficient suppliers to increase their market shares by lowering their prices. Schemes predicated upon uniform charges across providers preclude this process and thus are unlikely to change technical efficiency. For example, if charges across providers are uniform and consumers exhibit some price sensitivity, a policy of direct charges may reduce both the quantity demanded and, if countervailing forces do not overwhelm this initial reduction, the level of utilization.³ There will be no reason to expect changes in efficiency, however, because nothing has occurred to encourage individual practitioners to increase output per hour of their own time or substitute less costly auxiliaries for themselves or to encourage hospitals to speed up patient throughput.

While 'differential charges across providers' is a necessary condition for improvement in technical efficiency, it is not sufficient. Herein lies the second important distinction. The identity of the party determining the differential charges — provider or non-provider — may well be as important as the differential nature of the charges themselves. In theory, provider-determined differential charges allow sufficient scope for price-competitive behaviour to lead suppliers to more efficient production. However, a wide range of institutional mechanisms and rigidities can develop, as in the United States, to ensure that this potential is not realized. Restrictions on entry,

- 2 This assumes, of course, that the base of comparison is a universal first-dollar insurance plan. It is important to distinguish here between the notions of demand and utilization (Stoddart, 1975). While there may be an initial reduction in quantity demanded as represented by the number of initial patient/supplier contacts, secondary supplier-generated utilization effects make it difficult to predict a priori the total impact of a given direct charge on utilization.
- 3 If demand is completely inelastic (a fairly plausible assumption for some types of health care required to combat life-threatening situations) and patients modify neither their choice of provider nor the quantity of care demanded, then direct charges will serve only as a tax on those with health expenditures.

professional sanctions against advertising, publicly delegated professional self-regulation of quality defined to include patterns of practice, and the historical resistance of professional associations to innovative organizational modes such as prepaid group practice have all tended to suppress effective price competition. A non-provider determined option may therefore be necessary to encourage efficiency through differential charges. This does not imply that non-provider determined differential charges *guarantee* technical efficiency. But the likelihood of satisfying certain preconditions for technical efficiency may be maximized under certain forms of non-provider determined differential charges which we deal with in the following two chapters.

The present chapter analyses specific proposals for uniform charges. As indicated in Figure 2, these proposals may be formulated as either provider or non-provider determined charge schemes.

COINSURANCE

The simplest structure for a direct charge is fixed-percentage coinsurance, under which the patient is required to pay a certain percentage, 20 per cent, 50 per cent, or whatever, of all costs incurred on his behalf.⁴ Consideration is limited at this point to charges standardized across providers, which imply uniform binding fee schedules.⁵ Fees received by providers are assumed held constant. The administration of such a deterrent-charge system can be quite important in ensuring that proportionality is maintained. Both third-party-administered and provider-administered plans will be discussed. Within the latter, patients could be charged at point of service or periodically (e.g. monthly, quarterly, annually). Without digressing into a general discussion of analytic distinctions between provider-administered and third-party-administered coinsurance schemes (for that is dealt with later in this section),

- 4 A situation in which each individual provider could determine a rate of coinsurance for his or her entire patient population (so that intrapractice uniformity was maintained) but differential rates obtained across practices is definitionally a coinsurance structure. But it is a short step from there to differential inter- and intrapractice per centage charges, a topic taken up in chapter 4 in the discussion of extra-billing by physicians. In addition, the existence of rate differentiation across practices suggests that this is a special case within extra-billing rather than a question of uniform coinsurance charges.
- 5 Clearly a charge invariant across providers for a given service in conjunction with a fixed percentage coinsurance rate implies a uniform, binding fee schedule; otherwise actual out-of-pocket costs to consumers would vary with choice of provider.

FIGURE 2

Methods of increasing patient participation in payment for health care: charges uniform across providers

		Charges determined by	
		Non-provider	Provider
Level of charges across providers	Uniform	Coinsurance Deductibles Per-service charges Income- and income tax-linked proposals	Coinsurance Per-service charges
	Differential	D	C

it is worth noting one characteristic of the administrative function. A basic scale argument suggests that total administrative costs will likely be greater in the provider-administered plans than under third-party administration, although empirical verification would require an allocation of practice costs across functions (e.g. administration, delivery of care, continuing education). For the moment we abstract from this issue by assuming that direct reimbursement of physicians continues as at present in the Canadian system but that patients are billed periodically by the Ministry of Health for, say, 20 per cent of all their hospital and medical costs.

It is clear that *ceteris paribus* this would lower the cost of health services in the public budget by 20 per cent. This does not in itself have any direct impact on health costs but merely transfers wealth from citizens-as-patients to citizens-as-taxpayers (the obverse of a transfer of costs in the opposite direction). Such a transfer would be of dubious merit in terms of our general distributive criteria since tax burdens seem to be distributed more equitably over the population than medical and hospital costs.⁶ Moreover, it would

6 Manga (1978) found a generally positive relationship between family expenditure for medical services and income class for Ontario families, although adjusting for family size, location, and sociodemographic characteristics eliminated any significant dif-

reintroduce significant risks for part of the population, because the present technological and institutional focus of health care generates very large bills for some patients, and 20 per cent of a large number is still a large number. Therefore, if we assume no consumer, supplier, or system response to the introduction of a coinsurance charge, the apparent impacts on both wealth transfer and risk reduction would be perverse. The *ceteris paribus* assumption, of course, imposes inflexibility of patterns and levels of utilization, and as noted above uniform charges in general preclude any impact on production efficiency.

To reduce rather than shift the cost of health care, a coinsurance proposal would have to induce patients to use less care. For that to occur, four major assumptions must be satisfied. First, patients must be sensitive to prices in making decisions about health care use. Second, private insurance must not step into the gap left by the reduction in public funding. Third, providers of health care must not react to the new measure in such a manner as to offset the initial effect. Finally, as a practical matter patients must in fact comply with the coinsurance requirements and pay their bills. How plausible are these assumptions?

Ideally, one would wish for empirical studies identifying and measuring each of these responses. How much, if at all, will patient-initiated health system contacts fall in response to the introduction of a given direct charge? By how much will physicians then increase their service volume per patient-initiated episode? What proportion of the population will purchase supplementary private insurance to cover the direct charge, thus negating any incentive to utilization response? What proportion of the population will in fact be excused direct charges by reason of age, poverty, or simple refusal to pay the bills? The quantitative answers turn out to be remarkably difficult to ascertain.

The problem is that all these behavioural responses are interdependent. Data can be assembled showing differences in utilization under different insurance systems, with or without charges. But it is extremely difficult to partition utilization differences into patient-generated and provider-generated effects. The pattern of exclusion affects the size of utilization response; if the aged and low-income groups are excused direct charges, the average utiliza-

ferences in family expenditure across income classes. This result implies that expenditure as a percentage of income falls as income rises. Furthermore, Barer and Manga (1978) report that hospital utilization (as represented by days of stay) is significantly and inversely related to income class. Moreover, within any income class, illness costs are distributed capriciously and independently of principles of horizontal equity.

tion response is less, because these groups appear most price-sensitive. Who purchases supplementary insurance, and how much, will depend on the magnitude of the direct charge. Moreover, data from Canadian experience before universal hospital or medical coverage arrived may not adequately represent how people would respond to direct charges introduced into established universal systems. Subject to these qualifications, however, one can draw certain conclusions from the available evidence. In doing so it is useful to separate hospitals and medical care because utilization patterns for these two services are determined differently.

Hospital services have been traditionally regarded as at best very weakly price-sensitive. Undoubtedly some direct charges would be high enough to discourage use. But the critical role of the physician in the admission, treatment, and discharge decisions implies a relatively limited role for consumer choice in response to price signals. Canadian experience tends to confirm this. An examination of trends in Canada before and after hospital insurance led Evans (1975) to conclude that there was no significant utilization response to the introduction of public insurance. If the reduction of out-of-pocket costs to zero did not trigger increases in hospital use, presumably the reintroduction of direct charges would not induce decreases. One could of course argue that private hospital insurance in Canada was already so widespread that utilization responses to reductions in out-of-pocket costs had already occurred before the public programs came in. But recent unpublished research by Irazuzta (1979) has shown no evidence of a utilization response to changes in out-of-pocket costs for hospital care in Canada over the whole period 1932-74, spanning the periods of no insurance, private insurance, and public insurance. One would not then expect a coinsurance charge now to generate a response. Confirming this, Horne (1978) and Beck and Horne (1978) have reported studies of the hospital direct charge of \$2.50 a day in Saskatchewan from 1968 to 1971. After analysis of detailed disease categories, types of patient, and levels of hospitalization, they found no evidence that this charge affected utilization. Hospital costs account for over two-thirds of insured health care costs. If they are not sensitive to direct charges, the potential impact of any such charge on total health costs is drastically reduced.

Moreover, even if direct charges did affect hospital care, it is in that area that supplementation by private insurance is more likely. Mueller and Piro (1976) report that in the United States in 1974, 60.4 per cent of the medicare population supplemented the (incomplete) public plan with private hospital coverage and 51.9 per cent did so with private surgical coverage. This supplementation appears to have increased again in 1975 (Mueller, 1977).

The NCHSR 'catastrophic illness' study (United States, 1978, 10) reports that 57 per cent of the aged (65+) and medicare-eligible *non-institutionalized* US population had supplementary insurance of some sort to cover the coinsurance and deductible features of medicare; the institutionalized proportion of the aged was presumably more likely to have this 'wraparound' private coverage. But the same study pointed out that 17 per cent of the medicare group also possessed supplementary medicaid coverage, so that only 26 per cent of the non-institutionalized aged population were left facing the deterrent effects built into the US medicare system. Presumably if one took account of the institutionalized aged, and then measured the share of total *expenditures*, rather than the share of total insured *population*, exposed to direct charges, the proportion would drop from about one-quarter to much less — perhaps 10 per cent! Coinsurance and deductible features in the US medicare program seem better adapted to providing a market for private insurance than to influencing health care utilization levels.

In Canada, universal first-dollar coverage goes back to the late 1950s, and data are correspondingly sketchy for the earlier period. Taylor (1957) reports an estimate that 67 per cent of the population of Ontario had private insurance coverage for hospital care in 1954. Of the remainder, a significant number were partially or fully supported by governments because they were unable to pay. Both supplementary private insurance and exemption from the direct charge negate any incentive to reduce utilization. Hospital use by the elderly or the poor is significantly greater than for the average member of the population, and those groups were least likely to be able to purchase private insurance (which, as noted in chapter 1, is why the public programs were introduced). If these groups are exempt from direct charges⁷ and private supplementary insurance is permitted, few may be left to pay out-of-pocket costs.

This consideration undercuts any policy of deterrent charges involving high coinsurance rates. Evidence suggests that hospital patients do not respond to out-of-pocket costs as high as those currently experienced; \$100 a day out-of-pocket might well cut use significantly. But quite apart from the impact of such a policy on needed care, it would not be politically feasible to

7 In fact, the Treasurer of Ontario recently stated that any premium increase would not affect the current subsidization of the elderly (*Hamilton Spectator*, 12 April 1978), and there is no logical reason for the government to adopt a different position for coinsurance charges. The latest provincial budget includes an estimated \$210 million OHIP subsidy for 800,000 Ontario pensioners, and total premium assistance of over \$400 million (*Globe and Mail*, 13 April 1978).

ban private insurance in the face of such high direct charges. And if direct charges are insured privately, their deterrence effect is also nil.⁸

The empirical evidence on patient utilization response in medical care is less clear-cut. Pre-medicare utilization data are not in general available to serve as a baseline. Enterline et al. (1973) reported no over-all increase in utilization in Quebec after medicare was introduced, which again would imply little or no price sensitivity. On the other hand a series of studies of the Saskatchewan copayment experience by Beck (1971, 1974, 1976) and Beck and Horne (1978) does show some response in over-all utilization levels to the direct charge levied in 1968 of \$1.50 per office visit and \$2.00 per home or emergency visit. The charges are relatively small, but allowing for inflation over the last decade these would have been the equivalent of \$2.50

- 8 There is at present an active debate in the United States in the context of proposed national health insurance programs over the extent to which private supplementary or 'wraparound' insurance will fill in any gaps left by deductible or coinsurance features in a public program. Those in favour of such features for their presumed deterrence effect argue that such shallow (under the deductible) or wraparound coverage will not be purchased, because it is irrational to insure small risks; those against point to the level of supplementation which now exists in the US Medicare program (medical and hospital services for the over-sixty-five population).

The argument that supplementation will not be significant appears to rest primarily on theoretical grounds, predictions being based on the behaviour of hypothetical fully rational and fully informed consumers purchasing insurance on an individual contract basis at actuarially fair premiums plus a competitive loading charge (Keeler et al., 1974, 1977). The linkage between the behaviour of this hypothetical construct and the actual process by which insurance coverage is purchased is merely asserted or taken on faith. In fact as noted by Mueller and Piro (1976) and Mueller (1977) present private supplementation of US Medicare is extensive and growing, and is closely integrated with the deductible/coinsurance features of the public program. While supplementary 100 per cent first-dollar coverage for ambulatory care on the Canadian model appears to be unusual (Keeler et al., 1977), the really large expenditure categories of hospital and in-hospital medical services are extensively covered. Nor is it clear that the limited scope of supplementary first-dollar ambulatory coverage is due to an absence of demand for such coverage or to the private insurance traditions of the United States system not having offered such coverage in the past. In Canada, where all the population is now accustomed to such coverage and where private non-profit firms sold such coverage successfully in the pre-medicare period, 100 per cent supplementation might be much more prevalent.

As for the 'rationality' argument, there is ample evidence that individual and group purchasing behaviour does not conform to the hypothetical model. Eisner and Strotz (1961) showed long ago that the widespread individual purchase of airline travel insurance could not be consistent with rational risk-averse behaviour. Evans and Williamson (1978) argue that the purchase of group dental and pharmaceutical insurance is equally inexplicable in terms of rational models of consumer insurance

and \$3.50 in 1976, a significant proportion of the average cost per OHIP claim of \$14.40 (Ontario, 1976).

Depending on the mode of estimation, Beck measures a total impact on utilization of between 5 and 7 per cent. The impact is a greater 12.6 per cent on the copayable services (Beck and Horne, 1978, 24) than in total (5.7 per cent), but one cannot assume that applying the charge to all medical services would have increased the response significantly, because the non-copayable services were such things as surgical and diagnostic procedures, which are at the discretion of the physician. Moreover, the strongest utilization response (an estimated 18 per cent drop) occurred among the aged and low-income groups (Beck, 1974). If these groups were exempt from the direct charge, the utilization response of the remainder is close to zero. In fact, in Beck's (1971) analysis some middle-aged, middle-income groups appear actually to have *increased* their use of medical services after the direct charge was imposed. This result could be a statistical artifact; but it is also consistent with two alternative theoretical arguments. The drop in low-income-patient-initiated visits may have lowered the non-monetary costs of access (waiting times for appointments or in offices). Or, physicians may have recalled more frequently the patients who did initiate visits, and the net effect is then positive for middle- and higher-income patients for whom the deterrent effect is weakest. Either of these effects, however, undercuts the cost-reducing potential of the deterrent charge and emphasizes its effects in reallocating care from more to less price-sensitive users.

purchasing. Goldstein and Pauly (1976) explore the conditions under which group purchase of insurance can conform to individual preferences, and while they conclude it can do so, the necessary conditions are so extreme as to be beyond the capacity of any real-world labour market. Workers cannot really change jobs freely until they find one with just the right fringe-benefits package relative to salary! Krizay and Wilson (1974, 29) give a particularly striking example of 'irrational' group purchase of shallow or first-dollar insurance coverage at a premium almost as large as the maximum expected benefit. It appears that the dynamics and objectives of group behaviour in this market simply do not parallel individual rational decision making.

The response of defenders of the rational individual model to these observations is generally (e.g. Feldstein, 1977) that such 'irrational' overinsurance, while indeed widespread, is merely a rational response to the socially dysfunctional incentives embodied in the favourable tax treatment of employer-paid health insurance premiums. If the tax treatment were changed to remove deductibility of such premiums, much less (and more rational) insurance would be purchased and supplementation would not occur. Again, however, we are dealing with hypothetical responses based not on evidence but on the predictions of the rational individual model, and the applicability of this model to the group situation is taken on faith.

Apparently, therefore, direct charges to patients can reduce medical care utilization, though as Evans and Wolfson (1978) point out the countervailing response by Saskatchewan physicians generating additional servicing may have been muted by large fee increases during the same period as the direct charge. The Saskatchewan fee schedule rose sharply in 1968, and most economic interpretations of physicians' behaviour suggest that the resulting increase in net incomes would reduce their tendency to respond to any fall-off in patient visits by generating new service activity. The physician would thus benefit both from increases in income and from a reduction in workload. If the fee schedule had not been increased, physicians might have worked harder to maintain patient utilization levels. This interpretation can be supported with federal data on Saskatchewan physicians' fee levels and utilization patterns, as shown by Evans and Wolfson, but these data are not wholly consistent with the provincial data used by Beck. In any case, whatever utilization effects exist are small and are concentrated on the most vulnerable members of society. If these people are protected, utilization responses to direct charges disappear.

Once again, one might argue that a more energetic deterrent policy which imposed relatively high coinsurance rates (e.g. between 50 and 100 per cent) on those able to pay, while excluding the poor and aged, might moderate utilization. But as in the case of hospital care, the poor and/or aged segment of the population uses more than its share of medical care. Exclusion of the population over sixty-five or in the lowest-income groups (e.g. assisted premiums) would remove about one-quarter of medical care costs.⁹ The reduction in use among the remainder of the population would thus have to be proportionately larger to achieve any given impact on total use. Moreover, the high levels of direct charges presumably necessary to achieve such effects would again create incentives to private insurance. Prior to medicare in

9 The Ontario Ministry of Health (Ontario, n.d. [1978]) reported that 21.1 per cent of the province's insured population were receiving full premium assistance as of 31 March 1977. Of the 1,760,800 persons fully exempt from premium payment, close to 940,000 were in the 65-and-over exemption category. Others included provincial and municipal welfare recipients, pensioned veterans, and low-income individuals and families. An additional 0.1 per cent of the population - those individuals and families with incomes classified as low but above the full assistance cut-off - received partial assistance. Of the over \$700 million fee-for-service payments made by OHIP to practitioners for services rendered in the period 1 April 1976 to 31 March 1977, \$115.7 million, or 16.3 per cent, was paid on behalf of the 65-and-over group, and all premium-assisted categories accounted for 24.6 per cent of total payments. It is in the hospital sector, rather than the medical services sector, that the aged segment

Canada, about half the population had medical coverage. In Ontario this reached 60 per cent (Canada, 1963), and about half of these had first-dollar comprehensive coverage. Those who did not were likely to be towards either the top or the bottom of the income scale, usually the latter. Thus, for medical care as for hospital care, a direct-charge policy based on high co-insurance rates could be expected to lead much of the deterrent effect to slip away into private coverage. Only those ineligible either for private insurance or for public exemption would be left to be deterred.

Finally, of course, the mode of collection of the coinsurance charge we have assumed creates serious administrative problems. We have postulated that a provincial government agency bills patients after the care has been received for a uniform percentage of the bills paid on their behalf. The administrative (and political!) costs of collection are likely to be high, especially from those with severe illnesses. Uncollectibles were a serious problem for private physicians before medicare; they would be even more so in this plan. Nor is it clear that a post-receipt-of-care bill of unknown amount and probability of collection will serve as a significant deterrent to use. Deterrence and administrative feasibility seem to require that the coinsurance be collected at the time of service by the provider. Then, however, the scheme ceases to be a uniform-charge program and becomes one in which physicians *de facto* determine who shall pay and who shall not. The cost-controlling effects of such a scheme are diluted by all the factors just referred to and by a few more to be noted below.

The Canadian evidence suggests that utilization effects from coinsurance charges are generally weak or negative, but there exists significant support for their retention in the United States. As Hall (1966, 1974) has pointed out, much of this is ideological hangover from the private insurance industry. Nevertheless there have been numerous studies of the impact of direct charges on utilization, some of which appear to show significant effects, and these tend to be quoted by supporters of such charges. A comprehensive review

of the population takes a far more dominant share of total expenditure. For example, in 1973 Canadian males aged 65-9 stayed in hospital an average of 5.2 days and that figure grew steadily with age to 20.2 days for the 85-and-over male category. The comparable female range was 4.4 days (65-9 years) to 18.7 days (85+ years). In contrast, the average for all males was 1.7 days, for females 2.1 days (Boulet and Grenier, 1978, 89). Thus, whereas Ontario's aged (65+) incurred per-capita medical costs of less than twice the provincial average (*ibid*, 85) in 1974-5, their per-capita hospital costs were considerably higher. Gross (1978) estimates that the 65-and-over age group accounted for approximately 38 per cent of all institutional (hospital plus residential) and practitioner care.

of all the studies of American data on utilization response is far beyond the ambitions of this report. A great many such studies exist, and they yield widely divergent results. Their differences can be traced in part to differences in the quality and form of the data base, as well as in statistical methodology, the scope of population and services covered, and the formulation of the question being tested.¹⁰ Nevertheless some general conclusions can be drawn.

Most American studies have focused on the utilization of medical care and the extent to which full first-dollar medical insurance coverage could lead to an increase in utilization above that associated with the present partial coverage. This is of course the other side of the direct-charge issue. The divergence of results is very wide; Newhouse, Phelps, and Schwartz (1974) estimate that insurance at present covers 40 per cent of bills for physician office visits. They believe that an increase to 100 per cent insured coverage could generate a very large increase in utilization, approximately 75 per cent, and presumably in costs as well. Even if expansion of coverage were to 75 per cent, i.e. a 25 per cent coinsurance charge per visit or a \$2 to \$3 fee per visit for those currently having less than 75 per cent coverage, and no charge for those already covered to that extent, they believe utilization would rise by 30 per cent. These estimates imply that a 60 per cent charge in the present Canadian context would lower use by $1-(100/175)$, or about 43 per cent and a 25 per cent charge by $1-(130/175)$, or about 25 per cent. It should be noted, however, that these results refer to physician *office visits*, a very different thing from medical care use. First, the service content of a visit (and its cost) can vary a great deal; second, a high proportion of medical care use (by the most ill patients) takes place outside the office setting – in the hospital or intensive-care unit. Thus the Newhouse et al. estimates focus on the most price sensitive subcomponent of services.

Even for this subcomponent, however, Salkever (n.d. [1975]) reports an analysis of consumer demand for entry to the health care system, where price effects on utilization should be greatest, and concludes: 'Our results ... suggest there is little basis for the fear (expressed by economists and others) that more complete ambulatory coverage will lead to a medical care system

10 What is commonly estimated is a 'utilization elasticity', reflecting the percentage response in utilization to a percentage change in price, rather than the customary demand elasticity. The distinction is made necessary by the integral role of the supplier in influencing utilization, which is discussed later in this section. It is important here because 'demand elasticity' embodies only exogenous consumer reaction to price changes, whereas available empirical evidence is rarely able to disaggregate consumer from supplier reactions. A comprehensive discussion of this particular issue is Stoddart (1975).

clogged with persons seeking treatment for trivial illnesses' (ibid, F-6). His finding that out-of-pocket costs have no significant influence on decisions to seek care for particular symptoms of illness is very tentative — as are all the US studies — but is consistent with the Canadian experience that universal first-dollar medical insurance did not appear to trigger a flood of new utilization. Prices clearly rose, and identification of utilization changes is difficult owing to lack of before-and-after data; but surely if an increase in office visits of the order of magnitude of 75 per cent (Newhouse et al.) had occurred, someone would have noticed!

Feldstein (1974) surveys several US studies and notes earlier findings of perverse or non-existent price effects on medical care utilization in his own work (Feldstein, 1970) and in that of others (Fuchs and Kramer, 1972). But he concludes (almost, it seems, with relief, 406) that a study by Scitovsky and Snyder (1972) 'establishes that coinsurance (and presumably any change in net price) can have a substantial and significant effect in a setting *where the quantity is determined by demand*' (our emphasis).

The last phrase begs the question. As Stoddart (1975) emphasizes in a review of the medical utilization literature which goes beyond the economists' models to look at social and psychological explanations as well, demand in the economists' narrowly defined sense of informed choice under price and income constraints is only a part, perhaps a minor part, of the phenomenon of utilization. Only under very special conditions can observed utilization be equated with demand. Even in these circumstances, as Phelps and Newhouse (1973) and Newhouse and Phelps (1974) have shown, a number of econometric problems in estimating utilization response to price have led to upward-biased estimates in several US studies they analysed.

These technical problems of data adequacy and estimation technique are of unquestioned importance, but they seem to us secondary to the major weakness of incomplete coverage and inadequacy of adjustment for supplier response that bedevils all US studies. Scitovsky and Snyder (1972), followed up by Scitovsky and McCall (1977), for example, found that a particular insured employee group reduced its per-capita rate of office visits by 24 per cent in response to a 25 per cent coinsurance rate. They also found, however, that the coinsurance rate led lower-income non-professional members of this group to drop out of the covered group. What happened to global use by the whole group? Nor can one safely generalize these results from a basically healthy population to cover the more complex medical services required by more acutely or chronically ill individuals. Finally, a partial study does not control for supplier response; extension of these results to the whole community implies that a 25 per cent coinsurance rate would drop physicians'

38 Controlling health care costs

gross receipts (and workloads) by 24 per cent. It is hard to believe that practice styles would remain stable in response to such reductions.

There are further issues in the debate. If reductions in utilization of medical services do occur, there is no guarantee that they will be concentrated on 'unnecessary' services. They may be; Scitovsky and Snyder found that the coinsurance charge reduced annual physical examinations significantly. They regarded this as a negative effect — reduction of preventive care — but in light of increasing questions about the efficacy of such examinations their reduction might be considered a benefit of a deterrent charge. As we point out below, however, an across-the-board charge is an inefficient way of achieving this result. If annual physical examinations are inefficacious for everyone or for a specified age group, they should not be covered by insurance at all. Maximum deterrence would be achieved by charging for inefficacious services at 100 per cent. As for services which have not been shown to be inefficacious one is on very doubtful ground in arguing that any reduction in utilization resulting from direct charges will have no effects on health. Even if charges do not affect over-all use, but merely redistribute access from low- to high-income groups, there may be negative health effects. One should recall the findings reported by Aday and Anderson (1975) that pediatric services in the United States are underutilized by children in low-income families and overutilized by those in high-income families, while the *total* utilization appears adequate.

A related problem was pointed out by Roemer et al. (1975) and confirmed with further econometric sophistication by Helms, Newhouse, and Phelps (1978). A program of direct charges to certain medicaid beneficiaries in California (the Medi-Cal experiment) had the desired effect of lowering ambulatory use, but raised hospital use and total program cost. Helms et al. estimate an 8 per cent drop in ambulatory visits in response to a charge of \$1 a visit, and a 17 per cent increase in hospital care, leading to a (statistically insignificant) 3 to 8 per cent increase in total costs of care for this population. The data as usual embody a number of problems, particularly in the comparability of study and control families. Moreover, the claims by Helms et al. for the superiority of their elaborate estimation technique are somewhat diluted not only by the fact that their results are qualitatively similar to those of Roemer et al. but also by the discovery that their model includes a *monotonic* age variable applied to a sample population with large numbers of children. Nevertheless, the general point stands; a direct charge may cut utilization and costs in one component of a social program while raising total costs.

There are several possible explanations for this finding, as Helms et al. observe. The locus of treatment may have shifted in response to the charge,

so that care previously provided in physicians' offices moved into the in-patient hospital setting to avoid the charge. If so, presumably health status and actual use were unaffected (except by the increased health risk resulting from exposure to in-patient wards) but costs rose. It is hard to believe that a one dollar visit charge could lead patients to request or physicians to generate a 17 per cent increase in hospital use.

Second, the result could be a statistical artifact resulting from the fact that the utilization response of patients charged the visit fee was compared to their previous use and to the experience of a 'control group' who were not charged but who were also not a properly matched control. In this case the experiment does not tell us much. The position taken by Helms et al. — that the 8 per cent ambulatory use decline makes 'clear that the own-price effect for ambulatory services is important' but the 17 per cent increase in hospital use may be a 'statistical artifact resulting from differential effects on the copay and non-copay groups' — is trying to have one's cake and eat it too. Either the matching is adequate or it is not; and if it is not, it cannot support the 'own-price effect'; that is, it cannot show that ambulatory care use responds significantly to direct charges.¹¹

Finally, there is the possibility suggested by Roemer et al. that deterrence of ambulatory use has indeed occurred. As a result patients have become more acutely ill before contacting the health care system and have been more likely to require hospitalization; the direct charge would then have both raised costs and lowered health status. They point to a drop in the rate of performance of immunizations, pap smears, and other preventive services as indirect support for this view. But it seems to us that the experimental period was far too short for negative preventive effects to show up, and in any case a 17 per cent hospital use increase implies remarkably effective prevention! Over-all, one cannot avoid concluding that, despite the econometric legerdemain, the effects found by Helms et al. are too large to be explained by either of the systematic effects mentioned, while the much smaller changes reported by Roemer et al. (about one-half of 1 per cent drop in ambulatory costs and a bit under 2 per cent increase in hospital costs for a total increase of just over 1 per cent) seem much more plausible for a \$1 direct charge.

11 The authors, all members of the RAND research group, appear to have been carried away by their desire to support the hypothesis of the significant influence and policy usefulness of direct charges, in which that group has invested an enormous research commitment. Certainly the hypothesis needs all the support it can get. We are reminded of the legendary Scottish preacher whose sermon was marginally annotated: 'Weak argument here. Speak more loudly'.

But the preventive issue may be important, because it ties in with the findings of Scitovsky/Snyder and Scitovsky/McCall that annual physical checkups were reduced significantly. Brian and Gibbens (1974), in their comprehensive presentation and discussion of the data from the Medi-Cal experiment, report a marked reduction in immunizations, pap smears, and prenatal care, which they refer to as the 'major negative finding' in evaluating the experiment (29). Unlike annual physical examinations, these services are generally considered efficacious preventive interventions. Questions have been raised about the frequency of performance of pap smears, which may reflect excessive utilization in some socioeconomic groups; but for a medicaid population the preventive efficacy of this service may still be significant at the margin. On the other hand Brian and Gibbens report very minor deterrent effects, 'border[ing] on the subliminal' (30), and their data on preventive care seem to establish only that a differential exists between copay and non-copay groups in the use of such services, *not* that the charge caused it.

The chief difficulty is that for many preventive services the efficacy issue is still open. The logical approach for public policy is to attempt to establish as rigorously as possible whether particular preventive interventions are or are not efficacious. This the Canadian federal government is trying to do in a number of cases. If a procedure is inefficacious it should not be insured at all. If it is efficacious, then one should not count it as a benefit of direct charges (as Brian and Gibbens appear to do, 56) that they reduce over-all utilization, because that reduction is concentrated on efficacious services! Either way, the coinsurance percentage (or flat charge on all office visits) seems to be an exceptionally ineffective way of achieving public objectives in the area of prevention.

The general conclusions from the US studies of medical care use do not differ greatly from those of Beck — direct charges *may* influence utilization in aggregate, but not by much, and many other factors are at work. Large apparent effects emerge only from narrowly defined services and populations, such as office visits by employee groups, and over-all effects are likely to be much smaller. As an example, recently reported estimates of 'demand elasticities' by Newhouse and Phelps (1976, 276-7) are under 10 per cent for those seeking any ambulatory medical care, implying that a 10 per cent increase in price has an impact of 1 per cent or less on the volume of care used, while the probability of seeking any care at all responds to the coinsurance rate with an elasticity (negative) of just over 10 per cent. These elasticities are of course undefined for changes away from zero price. They imply that if one started from a 50 per cent coinsurance rate (the patient pays half the visit cost) and moved to full coverage, so that coinsurance would rise 100 per cent, the

probability of any individual's using care would rise by about 11 per cent and the amount used by each user would rise by 7 to 8 per cent, for a total increase of 18 or 19 per cent. Then, reversing these, the introduction of a 50 per cent coinsurance charge would lower physicians' office visit use by $(100 \div 118 \text{ or } 119) = 14 \text{ to } 15$ per cent. Newhouse and Phelps also estimate a response of average price per visit to the degree of insurance coverage, but that effect can be ignored if physicians cannot raise prices independently.

This is only part of the story, however. The authors report estimates of physician office visit rates per capita for various combinations of coinsurance rate, income level, and self-reported health status, for those members of their sample with some visits to physicians (281). Coinsurance rates varying from zero (complete insurance) to 100 per cent (complete self-paying) shift utilization by 7 to 10 per cent for those in good health and 12 to 13 per cent for those in poor health. Income affects use by a maximum of 10.5 per cent (highest income to lowest) for those in good health and with no insurance coverage; for those in poor health the maximum influence of income falls to 5 per cent or less for all coverage levels. But the key variable determining physician use is, not surprisingly, health status. Of those using any care at all, people reporting their health as 'poor' have twice as many physician office visits as those reporting it as 'good'. Presumably people in poor health are also much more likely to use at least some care. The mean probability of the use of physician services was (a very low) 0.56; if half those in good health and all those in poor health saw a physician in the sample year, then visits per capita for the ill would be four times those for the healthy. Coinsurance charges per capita would be four times as high — for minimal deterrent effects.

However, the authors recomputed the visits-per-capita response for those persons having physician visits, in a way which doubles its response to coinsurance rates (283). This still leaves health status as the overwhelming influence on use; but the effect of coinsurance rates goes from elasticities of -0.18 to -0.27. This would increase the estimates of utilization reduction from a 50 per cent coinsurance charge to 21 to 22 per cent. The numbers, however, are decidedly shaky.

Using these estimates, then, an across-the-board 50 per cent patient coinsurance charge for medical care in Ontario might lower utilization by 14 to 22 per cent if no private supplementary insurance were bought, if everyone had to pay, and if physicians did not respond to the change by generating new demand. This utilization response is about two to three times that found by Beck, but the Saskatchewan visit charge of \$1.50 in 1968 was substantially less than 50 per cent of the visit fees. For a visit fee of \$10, say, Newhouse

and Phelps's estimates would imply a 14 to 15 per cent drop in use in response to a \$5 direct charge, or about 7 per cent in response to a \$2.50 charge, and if allowances are made for inflation since 1968 the estimates look rather similar.

The supplemental insurance and physician-influence qualifications become even more important in the case of hospital utilization and its response to direct charges. Feldstein (1974) reports a wide range of elasticity estimates of the response of total hospital patient days to out-of-pocket costs. At the low end he reports -0.2, or a 20 per cent drop in use in response to a 100 per cent price increase; for other studies the values range from -0.5 to -0.7. He believes that the lower absolute values are downward-biased, but the reasons he offers (capacity constraints and variation of quality of care with cost) are unconvincing. There is widespread evidence of extensive *excess* capacity in the US hospital system, so it is hard to see how utilization increases could be constrained by capacity. As for quality of care, students of health care are generally agreed that quality measures focusing on the number and mix of *inputs* rather than on therapeutic *outcomes* are seriously flawed. In any case, higher quality (in an outcome sense) at given prices should lower, not raise, demand.¹²

Feldstein does point out that 'lower estimates of the price elasticity are generally associated with nearly complete insurance coverage'. This is quite reasonable and indicates that at very high out-of-pocket cost levels hospital use may indeed respond to deterrent charges. These charges will presumably be much higher than anything seen before in Canada, which may be why we

12 This is, of course, in marked contrast to standard utility-maximizing theory where fully informed consumers prefer more to less and, by extension, higher-quality to lower-quality goods and services. Consumers in this framework will purchase more of a higher-quality service, given identical prices for all qualities of that particular service. In the health care sector, however, high-quality care rendered in response to a given episode of illness is associated with a reduced need for subsequent care in the resolution of that episode. Examples are not difficult to come by – patients of a high-quality surgeon will have lower post-operative infection rates than those of his colleagues; a high-quality internist will pin down an elusive diagnosis more quickly than his average-quality colleagues; etc. This notion of quality is real and would be reflected, as suggested above, in *lower* service demand for any given episode of illness.

Of course 'episodic demand' must be distinguished from lifetime demand. One could argue that the receipt of higher-quality care within a given illness episode will extend the patient's life and lead to increased future demand deriving from the onset of other illnesses. It may do so, or it may allow the patient to live relatively illness-free thereafter. But even if we grant that in the 'aggregate lifetime' context higher-quality care may increase demand, neither this phenomenon nor the intra-episode decline in demand can be rationalized within the simple economic framework.

have no evidence in Canada of insurance-induced utilization effects. But when out-of-pocket costs become high, of course, private insurance returns. Maintenance of the present prohibition on private insurance against hospital expenditures is unlikely to be politically feasible if consumers/patients are exposed to significant out-of-pocket costs.

The various US estimates of the effect of out-of-pocket charges on hospital use are subject to the same kinds of estimation problems as the medical care estimates. Horne (1978) provides a good critical summary of these studies and casts doubt on the reliability of the response estimates. Newhouse and Phelps (1976) produce estimates of the elasticity of hospital use with respect to out-of-pocket costs and coinsurance in the neighbourhood of -0.20, but the length-of-stay response is not significantly different from zero, and the admissions response is surprisingly high. The critical issue, however, is what hospital charges would do to private insurance.

American advocates of direct charges as a cost-control measure argue that it is 'irrational' for consumers to insure against small losses. The purpose of insurance is to spread large risks: hence major medical insurance should be preferred to first-dollar coverage, and people should not buy 'shallow' coverage. Yet purchases of such coverage are common. As noted above, in 1974 three-quarters of the non-institutionalized participants in the US medicare program (for those aged sixty-five and over) had bought supplementary private insurance to cover part or all of the deductible and co-insurance charges built into that program, or else had public medicaid supplementation because of their low incomes. For the institutionalized population, the supplementation rate is apparently higher, and the proportion is growing.

The response then is that the tax deductibility of group insurance premiums paid by the employer subsidizes the purchase of insurance relative to the purchase of care.¹³ If the tax law was changed, people would voluntarily purchase insurance with deductibles and coinsurance, and shallow first-dollar coverage would be competed off the market. Users of health care would then respond to prices paid out-of-pocket and reduce demand; utilization and prices would fall; the cost crisis would be over; and the millenium would arrive. The concern about private insurance covering any direct charges, coinsurance or otherwise, is misplaced — if the tax deductibility of employer-paid premiums is withdrawn, no one will buy such coverage. Presumably the same would hold in Canada.

13 See, most recently, Newhouse et al. (1978), but the basic idea goes back at least to Feldstein and Allison (1972) and Feldstein and Friedman (1974).

This 'magic bullet' theory seems superficial. A general discussion is given in Evans and Williamson (1978, chap. 7). First, even if one accepts this argument, the case for direct charges becomes not absolute but conditional upon specific changes in tax law. If the tax law is not, or cannot be, changed, the supplementary private insurance will be bought and deterrence fails. But assume the tax law is changed. It turns out that the argument is still a pure hypothesis (or, less politely, a guess) about how insurance buyers will behave. It can be supported by reference to the behaviour of a hypothetical, fully rational, and informed insurance buyer who, it can be shown, would not purchase first-dollar coverage but would prefer (at the implied price differential) policies with deductibles and/or coinsurance.

But the applicability of this model of behaviour to the purchase of insurance is itself pure hypothesis. Most insurance in the United States is bought by employee groups, and the assumption that employee representatives choose fringe benefit packages to maximize the after-tax earnings (money or equivalent) of group members is a very specific one (quite apart from aggregation issues) wanting further support which it has not received. It is certainly not a self-evident description of union behaviour. Nor is it self-evident that individual insurance buyers act like informed utility-maximizers. In fact there is evidence that they do not (Eisner and Strotz, 1961). One is left then with the feeling that the magic bullet tax reform hypothesis is more a statement of pious hope about how the world might work than a serious description of probable results. It would likely do no harm to test the proposition if US tax law could be changed, but the hypothesis seems rather flimsy support for a direct-charge-based health policy.

This somewhat extended discussion remains a rather superficial treatment of the literature on utilization and patient charges. But a complete discussion, if that were possible, would be a major monograph in itself and, more important, would not resolve the issue anyway. We offer below some quantitative estimates (guesses) as to the probable impact of utilization charges in Ontario. But we are fully aware of the uncertainties involved.

Pulling the estimates together, we believe that direct charges for hospital use — in the politically feasible range — are unlikely to have any significant impact on hospital use. Out-of-pocket costs of, say, \$100 a day would probably affect use, but after the exclusion of the poor and aged, and the re-emergence of private insurance, no one (or almost no one) will ever pay such charges! Evidence of inappropriate or unnecessary hospital use abounds, but direct charges are not an effective policy instrument to control it.

Hospital costs are two-thirds of insured health care costs in Ontario. Thus, at most, only one-third of costs would be subject to the potential influence of

direct charges. As noted above, about one-quarter of these costs are incurred by the aged or other premium-assisted persons. Assuming these groups will also be excused from direct charges, we are left with (one-third times three-quarters) one-quarter of total costs potentially sensitive to direct charges. The Newhouse/Phelps and Beck results quoted above can be massaged to suggest that a 50 per cent out-of-pocket coinsurance charge might lower use by anything from 14 to 22 per cent, which, if adjusted by one-quarter, implies that a 50 per cent charge would lower insured health care costs by $[0.25(0.14-0.22)]$ $3\frac{1}{2}$ to $5\frac{1}{2}$ per cent. Yet 50 per cent is a very large coinsurance rate. Even 100 per cent self-payment, after necessary exclusions, would cut use by 7 to 11 per cent. The impact is underwhelming.

Nor is it likely that relaxation of our assumption of zero hospital use response would make much difference. The share of hospital use accounted for by the low-income and aged groups is much greater than their share of physician use. Boulet and Grenier (1978) report that in 1976, Canada-wide, those over sixty-five were 8.6 per cent of the population but generated 14.4 per cent of physicians' expenditure services and 36.5 per cent of hospital expenditures. As observed in note 9, in Ontario total aged and other premium-assisted groups generated about 25 per cent of physicians' services, of which about 16 per cent were received by the 65+ age group. That same group incurred approximately 43 per cent of total hospital expenditures (Gross 1978). Even assuming that the remaining premium-assisted groups utilized no more than their share of hospital services, we still find that half or more of hospital care is supplied to persons who are likely to be excused deterrent charges on political grounds.

Moreover, hospital costs, unlike physicians' costs, do not move in direct proportion to use. They contain a large fixed component, due partly to medical education and partly to fixed capital. There are various estimates of this fixed-cost component, depending on time horizon, but 20 per cent would be a minimal estimate and 50 per cent not wholly indefensible. Thus, even if one accepted Newhouse and Phelps's (1976) estimates that a 100 per cent increase in insurance coverage (50 to 100 per cent of patient bills) would raise hospital use by 20 per cent, implying that the reintroduction of a 50 per cent coinsurance charge in a fully insured system would cut use by 16 to 17 per cent, we have to adjust these down by about half (at least) for the excluded population, to 8 or 9 per cent, then by 20 to 50 per cent to allow for fixed costs, until we are left with 4 to 7 per cent of hospital costs eliminated by the 50 per cent coinsurance rate. But hospital costs are only two-thirds of total insured health costs. So it appears that a 3 to 5 per cent reduction in insured health costs would be an outer estimate of the impact of a 50 per cent

coinsurance charge on hospital use, to be added to the 3½ to 5 per cent estimate of impact on medical care costs. The total estimated impact is thus 6½ to 10 per cent — hardly spectacular. And again, we should emphasize that this is an *outer* estimate; our most likely estimate is zero effect in hospitals and 3½ to 5 per cent in medical care. Such a policy seems unlikely to solve the health care ‘cost crisis’ — real or imaginary — despite the burden it would place on the ill.

The above calculations do not pretend to be in any way a definitive and precise statement of the quantitative impact of coinsurance. Nor are they based on a complete survey of empirical studies in the area, which would produce such a diversity of results as to be useless. What they do show very strongly, however, is that *very large* changes in the currently available empirical estimates would be necessary to provide quantitative cost-control justification for a coinsurance policy. It is hard to see how changes of this order of magnitude could be supported with any of the existing evidence.

Out of this general quantitative discussion we should emphasize two factors that would probably cause the already small impact of coinsurance on total expenditure to disappear entirely. First, we have buried the influence of providers in the estimates of aggregated utilization response to out-of-pocket payments. It must never be forgotten that total health care costs closely approximate total health care provider incomes. Any cut in utilization which does result from direct charges will (at fixed prices or per-diem charges) result in lower payments to physicians or hospitals or both. Physicians can respond in several ways. They can grin and bear lower incomes; or some can leave the industry (emigrate, change jobs), thereby restoring the incomes of the remainder. Unlike most workers, however, they also possess significant power to influence utilization directly. If fewer people come to the door, more work can be done on those who do. There is abundant empirical information to support the general impression that empty hospital beds draw patients and physicians find work to fill their time.¹⁴ This may be harder to do when patients are paying part or all of the bill, but Beck’s

14 Despite some resistance from more conventional economists, the concept of supplier-induced demand as a critical aspect of health care economics is now relatively widely accepted and increasingly showing up empirically. It is considered in a number of articles in Perlman (1974), employed in analysis of the physician service market by Evans (1972), Fuchs and Kramer (1972), and Wolfson (1975), emerges from the empirical work of Rivard and Contandriopoulos (1977) and Contandriopoulos and Lance (1976), and is integrated into a general discussion of health care economics by Fuchs (1974). Stoddart (1975) places the concept of supplier-induced demand in the context of models of the demand for medical care and isolates some of its impli-

(1971) study of copayment in Saskatchewan showed clearly that, for middle-income users and for certain services that paid the physician relatively highly such as comprehensive examinations, total utilization *increased* significantly under copayment. Over-all utilization fell, but certain segments rose. It is immaterial whether this is physician behaviour to restore incomes to external target levels or a genuine medical belief that more care is better and therefore when time is available one should do more for the patient¹⁵ — or even a question of patients responding to shorter queues in offices and less waiting time for appointments. The point is that the feedback effects are such that any reduction in utilization generates provider and/or patient behaviour tending to restore utilization and maintain the incomes and activity levels of providers. All attempts to 'cut costs' by 'deterrents' are implicitly efforts to lower provider incomes; to assume that providers will accept these cuts while possessing the power to prevent them by recommending additional services to patients is both a very strong and a naive assumption. The effects are direct in the case of medical services, but no less strong for being indirect in the case of hospital services. Hospitals are, after all, significant sources of capital and co-operant labour to physicians, even apart from the totally hospital-based specialties, and a reduction in hospital use lowers physicians' earning power just as does a drop in office visits. Hence the tendency for hospital use to follow hospital availability.

A second factor, though, is perhaps more important. There is much evidence of inappropriate and unnecessary use of both hospital and physi-

cations for empirical testing. Most recently, Evans and Wolfson (1978) summarize both the theoretical development of the concept and the empirical evidence relating to it, in a paper prepared for a major conference devoted exclusively to the phenomenon (National Bureau of Economic Research Conference on the Economics of Physician and Patient Behaviour, Stanford, California, 27-8 January 1978). Although most recent studies have focused on medical care, early work by Roemer (1961) suggested that a supply-induced demand process was operative for hospital care ('a built bed is a filled bed'). Since then, hospital utilization studies have consistently shown bed availability to be the primary determinant of utilization. In addition, Kushman and Scheffler (1975) have reported dentist pricing behaviour which is consistent with demand generation.

- 15 The 'more is better' technological imperative may take two mutually reinforcing forms. Providers doing everything possible to enhance patient health status are likely to prescribe more services because they feel more cannot hurt and will likely help; but in addition there is an incentive to use the latest technology because it is part of the 'what is possible' frontier. The fact that it is also more costly and may be personally relatively lucrative is of secondary importance from this perspective.

cians' services.¹⁶ There is some evidence that direct charges can reduce the over-all utilization of physicians' services, though as described above the countervailing forces of private insurance, public support in the form of exemptions for the elderly and the indigent, and provider influence over utilization will make these aggregate effects very small indeed. There is *no* evidence of any kind suggesting that the selective utilization reductions which result from coinsurance are in fact those involving frivolous or unnecessary services. Most traditional economic arguments for direct charges build into their basic assumptions the existence of full consumer information and rationality, from which it follows that price-responsive consumers will give up the least efficacious services first. But this is a theoretical assumption pure and simple, which seems to have little basis in the facts of health care delivery. Whatever leads people to have tonsils or breasts removed, for example, it is unlikely to be the low price of the operation!

In fact it is extremely difficult to find evidence for the consumer-abuse hypothesis upon which much of the direct charges argument rests. Surveys of physicians by Enterline et al. (1973) and Wolfson and Tuohy (1979) found physicians' estimates of the magnitude of patient abuse to be minimal. Furthermore, the Wolfson and Solari (1976) patient utilization study found that high utilizers were in fact patients with more complex or severe conditions. (Claude Castonguay, the former Quebec social services minister, has recently suggested that the most effective avenue for reducing 'unnecessary' utilization, however initiated or generated, is dialogue between the third-party payer and the providers themselves, through their professional asso-

16 As noted in chapter 1, note 7, 'inappropriate and unnecessary' in this context does not mean that different patients might disagree about the usefulness of certain services, depending on their basic values. Nor does it mean that services which do not prolong life are not necessary; improvements in life quality are surely also of value. Rather, the statement reflects a widespread literature (see also chapter 1, note 13) on health services that do no one any good and have not been shown to have any positive effects at all. It is true, of course, that as the definition of positive effects is extended beyond objective measures of health status to include the psychosocial benefits of 'being cared for' and 'talking to someone', the boundary between necessary and unnecessary services, at least in primary care, becomes vague. (Presumably one does not remove healthy organs to provide psychosocial benefits.) But as noted above, the more broadly one defines potential benefits, the harder it is to justify highly trained personnel and technically intensive care as an efficient response to such needs. In the extreme it is probably true that most of us as citizens would be unwilling to extend socially supported health care to the point of providing whatever makes anyone happy. The exact boundary between health status and happiness may be fuzzy, but they are different concepts nonetheless.

ciations.¹⁷ Examples of successful dialogues in Quebec and Saskatchewan are noted in Evans and Wolfson, 1978.)

The outcome of direct charges is that people may refuse recommended services they cannot afford, whatever the degree of 'necessity'. Thus the utilization effects of coinsurance, such as they are, tend to hit the poor most heavily. Since the 'needs' of low-income people will go unmet along with their frivolous demands, and since governments cannot distinguish between the two any more than patients can, it will almost inevitably follow that any group whose use is seriously inhibited by direct charges will receive some sort of public subsidy. Only those whose use does not fall, or falls very little, will be left fully exposed to such charges, which will serve only to transfer income to citizens-as-taxpayers from citizens-as-patients or buyers of private insurance.

Finally, if, despite all the adjustments and qualifications above, utilization and provider incomes actually do fall in response to deterrents, the fee bargaining by primary providers is likely to stiffen considerably. The scare stories circulating in the news during 1977-8 about doctors emigrating from Ontario and their efforts (noted in chapter 1) to convince patients and the media of the justice of higher fees are examples of attempts by physicians to loosen up public bargaining and increase their incomes through higher negotiated fee schedules and/or additional charges to patients. Either route is guaranteed to increase medical costs.

In terms of the objectives of this report, the pattern we have analysed (of government billing health care users periodically for a certain amount of incurred costs) is the simplest to study. A more operationally plausible form of coinsurance, however, would entail physicians collecting and keeping the coinsurance component. Government might negotiate a new fee schedule with physicians, or allow larger per-diems to hospitals, with the proviso that some percentage of this was to be collected from patients. This administrative variant would be placed within quadrant A of Figure 1. Alternatively, medical or hospital associations might independently promulgate a uniform percentage to be added to a government-determined fee schedule or reimbursement rate, a scheme that would fall within quadrant B. Either would be similar to the third-party administered plan, embodying all of the weaknesses of that simpler case as well as a few of its own. In particular, either is open to direct abuse by the physician in two ways that the first approach is not. First, physicians can bill for more expensive services, then forgo the

17 Testimony before the Select Committee of the Ontario Legislature on Health Care Financing and Costs, 8 July 1978, Toronto: Queen's Printer.

deterrent. There is a suggestion in Beck's (1971) data that this happened in Saskatchewan. Complete examinations rose, regional examinations fell, and anecdotal evidence suggested that not all physicians collected the point-of-service charge. Second, since coinsurance charges would apply at time of service, the physician would have an incentive to bill for services which might later be disallowed by the insurance plan if he need not return the patient's percentage. This transfer of risk for inappropriate servicing from provider to patient would reach its most extreme form in a setting where the patient is billed at point-of-service for the full cost of all care received and is subsequently reimbursed by the insurer. If a service was deemed inappropriate, the patient would again be at risk in the absence of any legal recourse. It is not surprising, then, that schemes in which patients are reimbursed by insurers are popular among providers ('Doctors want patients to pay them directly', *Hamilton Spectator*, 23 June 1977, 2).

A second weakness of the provider-administered variant derives from the function of administration itself. Bill-collecting and the associated accounting impose real costs on the physician (be they own-time or other practice costs). Once these costs have been incurred, the additional step of opting out and extra-billing is not large.¹⁸ At this point, total (as opposed to public) medical care costs become essentially uncontrolled. There is a redistribution of risk-bearing from the taxpayer to the patient, and the situation degenerates into one of individual provider control over the coinsurance amount, a topic considered in more detail in the following chapter.

One final drawback in the provider-administered option concerns the very serious bargaining issue of the level at which total fees or hospital budgets are set. We have assumed in this discussion that the total fee or budget received by the provider is independent of how that amount is split between patient and insurance plan. Once the provider is the direct recipient of the coinsurance amount, however, the bilateral bargaining relation is broken, and the possibility exists that provincial bargainers will be more likely to trade off a higher total charge (and higher over-all medical and hospital costs) to achieve a lower public component, with the difference made up by the coinsurance charge. In effect, then, some of the third-party/provider confrontation is shifted to provider-patient bargaining. The short-run gains of such a policy (to both provider and insurer) are abundantly clear, and the

18 In Ontario, for example, physician participation in OHIP is not compulsory. Those physicians not participating are designated 'opted-out' physicians. A physician must opt out in order to bill a patient directly, whether at or above the OHIP benefit schedule. Both opting-out and extra-billing are considered in more detail in chapter 4.

long-run losses (e.g. higher total costs, political ramifications, welfare burden), though more serious, are less obvious and, most important, less immediate. Trading off these longer-run unpleasanties for short-run gain is a clear case of not dealing with the important issues today if they can be put off for someone else to handle tomorrow. Thus, a provider-collected coinsurance feature would have a tendency to expand and to lead to higher total costs, just as it appears to have done in the US medicare system. If the uniform coinsurance rate is provider-determined as well as provider-collected, there will of course be no public-negotiation constraint on total cost escalation.

The proportionate coinsurance system (in any administrative guise) seems therefore to have little to recommend it in terms of its ability to facilitate cost control through inducing consumer or provider responses on the price or utilization side. The incentives inherent in such schemes will tend to encourage neither reduced unit costs nor significant shifts in patterns or levels of utilization. In the unlikely event that significant utilization reductions do result, there is no evidence that the deterred visits will be those least medically necessary. In any case, reactions by providers to lowered patient-initiated utilization (increased provider-generated utilization and/or higher prices) will be in a direction opposite to what little initial utilization response occurs.

In addition, and depending on the nature of any exemptions, coinsurance will increase the risk levels to which some consumers of care are exposed (relative to a 100 per cent first-dollar insurance system). And, with respect to the wealth redistribution objective, any transfer will in general be regressive, from the ill to taxpayers as a whole. These latter two perverse effects (increased risk and inequitable transfers) will tend to reinforce the elimination of any initial utilization response by encouraging the purchase of supplemental private insurance and government subsidies to the poor, the aged, and other high users. Finally, there will be no significant impact on the efficiency of operation of the supply side. Since coinsurance charges are uniform across all providers, the 'efficient' supplier (if such could be identified) cannot expand his clientele by adjusting prices. This observation is no less true for any other uniform direct charge and will thus be a common theme in the rest of this chapter.

DEDUCTIBLES

One alternative to coinsurance is the deductible charge. Rather than making the patient liable for some percentage of all bills, the deductible requires that he pay 100 per cent of all bills in a given period up to some maximum

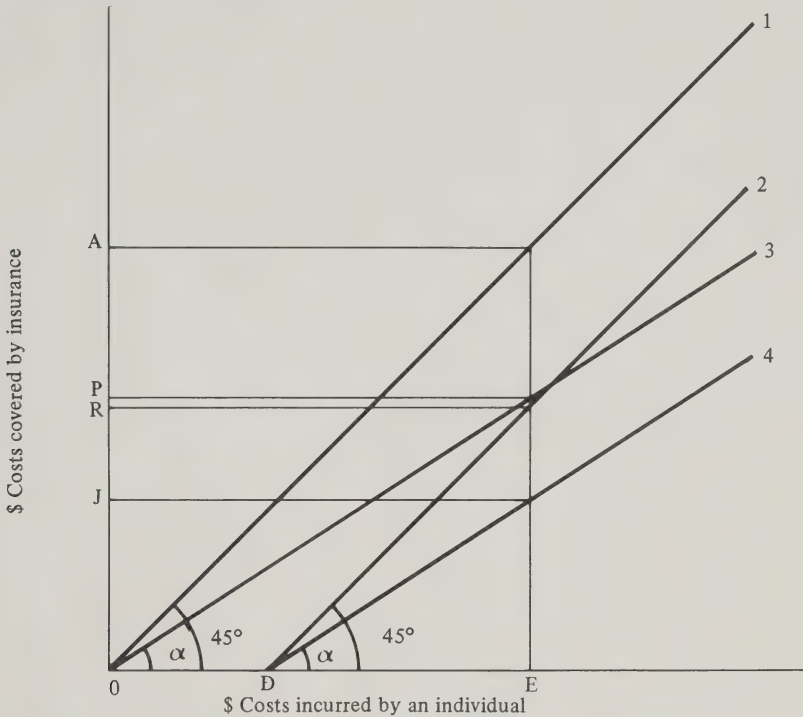
amount, beyond which insurance benefits are paid. A common feature of most other forms of insurance, such as automobile and house insurance, the deductible approach is also employed by the British Columbia and Manitoba pharmacare programs, in which the covered individual or family pays the first \$50 in Manitoba and the first \$75 in British Columbia. While these particular plans have also added a coinsurance provision, making the beneficiary liable for 20 per cent of all costs above the deductible limit, such a provision is not a necessary feature of the deductible approach and of course is subject to the general criticisms of coinsurance expressed above. The difference between a deductible and both universal first-dollar coverage and coinsurance coverage appears in Figure 3.

A representative consumer incurring \$OE in health care costs would pay \$OD=\$AR under a straight deductible plan. The plan would cover the remaining \$DE=\$OR. The percentage coinsurance scheme would leave the consumer with direct payments in the amount \$AP, with \$OP being the costs covered by the insurance. Under the 'deductible plus coinsurance' plan illustrated as scheme 4, the out-of-pocket costs would be \$AJ. For $E = \$300$, a deductible of \$115 and a coinsurance rate of $33\frac{1}{3}$ per cent ($\alpha \approx 30^\circ$), the respective direct charges would amount to \$115 for scheme 2, \$100 for scheme 3, and \$176.67 for scheme 4.

The deductible approach has potential value for services where utilization is relatively common and costs for most users are low. Most people purchase some prescription drugs in a given year, for example, but very few spend large amounts on them. Most people with large drug bills associated with severe illness receive drugs without charge as part of a hospitalization episode. Effectively, then, the deductible drops most of the population out of the insured group and focuses the insurance function on the very small part of the population with a risk of high expenditure. In doing so, it also eliminates the high administrative cost component associated with handling a large number of small claims. Although the plan is called 'universal', most of the population will pay their own drug bills and will have little or no chance of receiving benefits because their probability of over-deductible use is so low. This is of course entirely consistent with insurance principles. There is no gain from 'insuring' small and relatively predictable expenses. On the other hand hospital expenses are relatively infrequent, unpredictable, and costly. Anyone who has an episode of hospital care will generate relatively large expenses. In this environment it makes little sense to use a deductible. A low deductible would be exceeded by every patient admitted; a high one would impoverish users and would be undermined by a combination of private insurance for those who could afford it and government subsidy for those who could not.

FIGURE 3

Comparison of consumer participation schemes in relation to out-of-pocket expenses of patients



Note: Full first-dollar coverage (01); Deductible plan (0D2), deductible portion (0D); Percentage coinsurance (03); Deductible-plus-coinsurance plan (0D4), deductible portion (0D).

The chief strength of a deductible plan, therefore, is not utilization deterrence¹⁹ but rather division of the population into groups, one with and the other without risk of high expenditure levels. If in addition one seeks to

¹⁹ Given the nature of the product 'hospital care', and the fact that the decision to admit usually rests with the attending physician, it is difficult to see how such a fee is likely to deter anyway in this demand-inelastic case. Evidence cited earlier (Horne, 1978; Beck and Horne, 1978) suggested that small deterrent fees had no effect on utilization, although there is undoubtedly some deductible level high enough to deter use. For such high deductible levels, however, barriers to access to necessary care and the re-emergence and spread of supplementary private insurance become critical issues.

lower utilization levels by the deductible approach, all the problems of countervailing behaviour by providers and exemption of significant proportions of the population, particularly the elderly and the poor, come into play. A deductible applied to hospital care, for example, would amount principally to a tax which transferred funds in the amount of the deductible from each person admitted to hospital to the general provincial revenues. Merely from observing the very pronounced age-related pattern of hospital use one would expect such a tax to be highly regressive, shifting wealth from lower- to higher-income groups. The elderly members of the population are both relatively low-income and very high users of hospital care; Boulet and Grenier (1978) estimate that per-capita hospital costs (in 1974 dollars) for the elderly ran from \$438.45 for those aged 65-9 to \$1364.60 for those over eighty-five (males and females averaged with equal weight) compared with \$173.80 for the general population (the figure for the population under sixty-five would clearly be even lower). This strong indirect evidence of regressivity is confirmed by the results of a household survey reported in Barer and Manga (1978), which demonstrated that hospital expenditures are indeed higher for lower-income individuals. Both days of stay per person and average length of stay per hospital admission were significantly and inversely related to income class. Those in the \$0-3999 income class were hospitalized an average of 6.4 days a year (14.2 days for the subsample composed only of those persons with one or more hospital episodes), and this fell to 1.3 days for the \$20,000+ income class (2.5 days for those hospitalized). The average length of stay was approximately 13.5 days for each admission of a person in the lowest income class, as opposed to 5.8 days for the \$20,000+ class. Applying estimates of hospital-specific in-patient per-diems to these utilization rates, the authors reported in-patient care expenditure of around \$560 for each person in the \$0-3999 class, significantly higher than the \$130 estimate for the upper-income class. Again, the inverse relationship was significant ($p < 0.01$ in all cases).

If a deductible did deter any hospital admissions, hospitals and physicians could keep beds filled by raising lengths of stay for those who have been

Some recently published work in the United States, however, suggests that deductibles can in fact have a significant influence on health care expenditures (Newhouse et al. 1978). This study will be considered in more detail in chapter 4 in discussing major-risk insurance, because it is based on a regime in which providers set their own charges while insurers reimburse a proportion of expenses above a dollar limit per time period. The results of that study do not appear to justify modification of the present discussion.

admitted anyway and who are not, at the margin, paying for their extra care. Similarly, physicians can in good conscience 'do more for their really ill patients', that is, for those whose care costs will exceed a medical expense deductible, if such a charge were actually to reduce the number of patients who initiate a care episode. Whether either of these adjustments would be beneficial or harmful to general health levels depends on whether the deterred admissions/ visits are the least medically necessary ones. If that were so, it would imply that patients can accurately judge this degree of necessity. No evidence has ever been adduced to support this assumption; but even if it were true, the providers of care still have significant scope for increasing utilization at the margin among the population of users who have exceeded the deductible.

Another justification for the deductible in pharmaceutical 'insurance', apart from the fact that most of the population are not at risk and do not need coverage, is that by maintaining the exposure of most of the population to market prices it leaves open the possibility of using market forces to induce significant improvements in the organization and efficiency of the drug delivery system (see Evans and Williamson, 1978, chap. 3). Such market forces can only be used effectively when the product or service being supplied (in this case the drug dispensing process, not the drug itself) can be standardized so that the consumer can judge its value and relate this to price. Consumers are not in general competent to choose among different antibiotics or tranquillizers, so there is no point in charging prices for these drugs that would 'guide' such a choice (much less the choice of whether or not to purchase the drug at all after a physician has prescribed it). But one might quite reasonably expect consumers to choose between different dispensing outlets on the basis of dispensing charges. This example illustrates again the significant differences between uniform and differential charges, to which we return in detail below; however, it is worth noting that deductibles, like all direct charges, may play a more useful role in differential- than in uniform-charge systems because of the possibility of influencing efficiency in the former.

In terms of our policy criteria, and compared to a universal first-dollar coverage system financed by premiums and/or other taxes, a deductible approach would increase the financial risk to an individual of both small and frequent expenses like pharmaceuticals and large infrequent expenses like hospitalization. It is also less desirable on wealth transfer grounds, at least for levels of expenditure below the deductible ceiling, since it removes the low-to-high-user transfers implicit in the current system. Furthermore, the deterrence effects of deductibles in lowering overall utilization and costs, and the likelihood of deductibles encouraging cost-effective patterns of utiliza-

tion, will be small for the same reasons as with coinsurance. If consumers are not price-sensitive, or if private insurance insulates against the price effects of deductibles, or if provider reaction undermines deductible-induced demand reductions, then deductibles will have little impact on levels of utilization or costs. Even if all these conditions were to hold and utilization did fall in response to the deductible, there is no reason to assume that the most price-sensitive services will be the least necessary and/or efficacious. Moreover, if providers collude to set uniform charges and regulate patterns of practice, deductibles, like coinsurance, will prove useless as a stimulus to technical efficiency.

This is the point which advocates of 'major-risk insurance'²⁰ routinely miss: a deductible for health services where competitive pricing does not function, either because of supplier collusion or because of inherent service characteristics such as the inability of the consumer to judge need, serves no useful function. If consumers cannot respond to relative prices in their choice of provider, then in the case of unchanged utilization the deductible is just a form of tax on the ill. One can develop highly sophisticated methods of reallocating the burden of this tax across income classes by manipulating income tax schedules and establishing related short-term loan funds; but regardless of its incidence by income class a deductible plan without workable competition in the associated health care market is (compared with first-dollar coverage) simply a means of facilitating a transfer of wealth from the ill to the well.

PER-SERVICE CHARGES

This approach identifies certain types of services and requires the patient to pay a flat charge which may or may not be related to the actual cost of providing the service. Examples of per-service charges would be the \$4 per-patient-day charge for hospital space in British Columbia, or the utilization ('deterrent') fee of \$1.50 per physician office visit introduced in Saskatchewan in 1968 and later discontinued.

Flat-rate charges on particular services have the same transfer effects as deductibles, but in certain circumstances that may be exactly what is desired. A per-diem charge for occupants of chronic-care beds or for long-term care facilities for the elderly may be set approximately equal to the old-age

20 The principle of 'major-risk' insurance is that no one ought to be at risk of financially devastating health care costs. Thus, such plans embody a ceiling on individual out-of-pocket expenditures. The concept is considered in more detail in chapter 4.

pension, or a percentage of it, thereby leaving pensioners some pocket money. In this case the purpose of the charge is not to deter use or to hold down system costs. Rather it is to reflect the fact that for disabled elderly persons, long-term care is an alternative to some other domiciliary arrangements for which they would have to pay out of their own resources or public pensions. It is distributionally unjust for the *heirs* of an elderly person to profit by institutionalizing that person at public expense and then banking the public pension or earnings from the elderly person's capital (if there is any) which were intended to pay for that maintenance until, in due course, the cumulated pension and other earnings become part of the estate. It seems in this context wholly appropriate for a maintenance charge in long-term care to be set high enough to absorb the resources intended to cover that maintenance, particularly the public pension component. While such a maintenance charge increases financial risk compared to a universal first-dollar coverage plan, this increase in risk is entirely consistent with the intent of the wealth transfer criterion. Although the objective of maintenance charges is not deterrence, compared to a universal first-dollar coverage scheme such charges would create a strong incentive for home care wherever possible. Whether or not such charges would result in more efficient delivery of long-term care would depend upon the relative cost and effectiveness of publicly (and privately) provided home care services and the same services provided within an institutional setting. Unfortunately, a paucity of empirical evidence precludes a more definitive assessment of the efficiency question.

It is extremely sloppy thinking, however, to expand the maintenance concept to the acute-care hospital scene and suggest that patients be made responsible for 'hotel' or room and board charges on a per-diem basis. Apart from the obvious incentives this would create for hospital bookkeeping (with government blessing) to shift as many costs as possible into the hotel component, it confuses the nature of the alternatives to care. In long-term care, maintenance in the institution is an alternative to self-maintenance or maintenance at family expense. The institutionalized individual does not incur these alternative expenses, and the savings might reasonably be offset by an institutional charge. But the individual in an acute-care bed for three days to three weeks (which covers almost all acute-care users) does not sublet his or her apartment or room in the family home for that period. Most of the maintenance costs for an individual in acute care are fixed, in the short run, and do not disappear during institutionalization, just as they do not disappear for a family on summer vacation. Thus the hospital costs are additive to, not substitutes for, self-maintenance costs, and there is no saving for the per-diem 'hotel' charge to offset. This is even more obvious if ward nursing care is

smuggled into the 'hotel' costs, since very few users of acute-care hospitals have regular three-shift private-duty nurses at home whose services can be 'saved' during hospitalization! The only conceivable substitution argument would apply to food costs, which the hospitalized patient would otherwise incur at home, but which are a small share of in-patient per-diem acute care costs (Barer and Evans, 1978). Since even these would have to be net of special therapeutic diet and handling costs, all the patient really 'saves' is raw food costs, which would result in a trivial per-diem charge.

But if the per-diem charge is not an offset to other expenses the patient saves by virtue of being in the institution, its justification on equity or wealth transfer grounds collapses. On risk-spreading grounds it is unattractive, since it penalizes those with long stay and will tend to induce private insurance and the associated adverse selection problems. Its deterrence effects on utilization meet with the same criticisms as do coinsurance and deductibles. Apparently efficiency grounds afford the only potential justification for such a charge. Here one might argue, in a manner similar to the proposal by Newhouse and Taylor (1969) for variable cost insurance, that if different hospitals have different 'hotel' costs, and if smaller proportions of bills in higher-hotel cost hospitals are covered, patients might choose lower-cost hospitals.

Such an approach, however, requires two major behavioural assumptions. First, it assumes that patients exercise significant choice of hospitals and would be sensitive to price differentials in making this choice. If, in fact, patients choose physicians, and physicians hospitalize patients where they have admitting privileges, then patient sensitivity to price will be small. This will be true *a fortiori* if patients have perceptions of hospital 'quality' based on ethnic, religious, or community ties, which, valid or not, will tend to override price considerations unless the latter are large. Second, and perhaps more important, the variable-hotel-charge approach assumes that hospitals will in fact compete in this dimension. Collusion between hospitals may lead to uniform 'hotel' charges in each area, especially since any patient sensitivity to charge differentials will become immediately apparent to each administrator. An administrator who is losing patients may try to be more efficient; he may also attempt to reallocate costs from 'hotel' to other costs (met by insurance) until his charges are competitive. In any case, the problems of efficiency or inefficiency in hospitals appear to relate more to patterns of use of hospitalization in itself and of diagnostic and therapeutic procedures. Hotel costs seem rather peripheral. Consequently, prospects for engendering efficiency-enhancing price competition between hospitals through the introduction of per-service charges appear limited. Overall, although the selective

application of per-diem charges in long-term care to balance self-maintenance expenses saved by institutionalized patients may be a justifiable step, it is one unlikely to have much impact on current health cost levels. The significance of such a policy may grow in future decades when the aged component of the population begins to expand to absorb the post-war baby boom, but it remains a tax-transfer policy, a policy to redistribute health costs rather than to affect their level.

The scope for use of fixed per-service charges in other areas of health care appears equally limited. As far as influencing the level of utilization and expenditures is concerned, the direct charge per office/home/hospital visit or per prescription may indeed decrease demand to some degree. However, Beck's (1971, 1974) results cited earlier show that its total effects tend to be limited and heavily concentrated on lower-income users. Such direct charges thus fail on wealth transfer grounds, have little or no impact on the efficiency criterion, and raise serious questions about the appropriate distribution of care. Recall that Beck's demonstration that fixed point-of-service access charges swung utilization down significantly for low-income people and up for middle- and upper-income people²¹ is consistent with Aday and Anderson's (1975) report that 'expert' evaluations of needed pediatric care for children in the US found need and use to be approximately in balance — overuse by higher-income people balancing underuse by those in lower-income groups — in a system embodying significant direct charges to patients.

Although the general deterrence effect of per-service charges is thus likely to be small and of dubious impact on health, such charges may possess some utility for the limited range of health services entailing little or no health benefit. One might therefore reasonably argue that since annual physical checkups have yet to demonstrate any significant effect on health status, some per-service charge should be levied on them to discourage utilization.²²

21 For a comprehensive discussion and review of the Canadian empirical literature on patient cost sharing, see Beck and Horne (1978).

22 Sackett (1972), after discussing the validity of traditional methods employed in evaluating the clinical effectiveness of early detection programs and presenting selected empirical evidence on the effectiveness of health examinations, concludes that the majority of the screening and diagnostic procedures applied during periodic health examinations are being applied either in the absence of, or in direct contradiction to, evidence of their effectiveness. More recently, the Federal Task Force on the Periodic Health Examination has been systematically evaluating the procedures which comprise the examination and is expected to make public its findings within the next year.

If one adopts this stance, however, it is difficult to argue with the prescription of per-service charges equal to the full fee for inefficacious procedures. The extension of the per-service charge approach to what is effectively 'selective deinsurance' represents a policy direction demanding much closer analysis of its operational feasibility, economic impact, and requirements for clinical evidence. This examination is undertaken in chapter 5, in conjunction with other prospective policy directions involving non-provider determined differential charges.

But the possible kinds of uniform charges have not been exhausted. To this point, we have suggested on several occasions that specific exemptions (while they may be socially desirable) would tend to diminish any impact of uniform charges on utilization by reducing the population at risk, but we have given only passing mention to the operational problems of determining exemption eligibility. The machinery for that type of decision-making is already in place in the form of the income tax system in Canada (and the United States). In the following section, then, we consider both administrative and 'exemption status determination' roles for that system.

INCOME- AND INCOME TAX-LINKED OPTIONS

One of the policy prescriptions contained in the Ontario Economic Council's (1976) survey of the health care delivery system in Ontario was a scheme in which a patient's direct financial responsibility for health care would be a function not only of costs incurred but of income class as well. More specifically, an income-related share of a patient's costs would be 'charged back' to the patient. The examples put forth by the Council and described in greater detail below are but a subset of potential linkages between use, income, and direct charges. A limited number of alternative options will also be considered in this section.

At the outset it may be noted that, since none of these schemes involves changes to the pricing or delivery structure for health care, they are uniform charges in the sense adopted earlier. Thus, while a patient's financial burden will vary according to income class, it will not be a function of the provider chosen. But all the uniform charge plans described earlier were seen to be regressive, or at best not obviously progressive. Excused consumers aside, coinsurance, deductible, and per-service charge plans all introduce a perverse change in the wealth transfers inherent in a first-dollar universal plan because of the distribution of illness incidence. In comparison, income-linked schemes such as those proposed by the Ontario Economic Council and (if we are willing to accept that the current Canadian income tax system is

more progressive than the uniform charges considered to this point) income tax-linked schemes will at least induce wealth transfers in the normatively correct direction. In that respect, they differ from all the other uniform charge approaches outlined earlier. Like any other charge system, however, they are horizontally inequitable in that the burden of health costs may be distributed unevenly within an income class.

The particular plans suggested by the Council involve adding to a patient's income tax liability an amount determined by applying a combination of deductibles and coinsurance to the total health care claims paid on his or her behalf. The deductible level and coinsurance rate would each be a function of income. A ceiling on direct financial liability for incurred health care costs would also be variable and related to income. Figure 4 sets out the essential features of three representative sample plans. Apart from obvious discontinuity problems inherent in these examples (a person earning \$14,999 will pay a \$100 deductible, while the marginal deductible cost of an additional \$1 in income is \$50), the gist of the plans is self-evident.²³ The first \$X of expenses (where X is a function of income) are the patient's responsibility, as is a share (also related to income) of all expenses over and above the deductible. Once the combination of deductible and coinsurance reaches the ceiling, no subsequent incurred costs are added to tax liability. While the Council's choice of examples involves a combination of deductibles and coinsurance, the basic nature of the scheme would be maintained with any of deductibles, coinsurance, or per-service charges in isolation or in alternative combinations as long as a ceiling was incorporated in each case and all features were functions of income.

It turns out that the wealth transfer distinction between these and the other uniform charge plans noted above is their only significant differentiating feature in terms of our four public policy objectives. Since no change is envisaged in the form of reimbursement of providers, no new incentives to technical efficiency are created. The combination of ceilings on total liability and low or non-existent charges for low-income groups will lead to exemption of a significant share of expenditures. For persons with expenditures that do

23 While this discontinuity is a rather blatant example of the 'notch' effect, it does serve to point out a general problem with any scheme requiring categorization of consumers. In the absence of a perfectly continuous income class scale, any collection of exemption categories will have built-in 'notches' which may impart perverse incentives to those near the lower margins in each category. The income tax system ameliorates the notch effect because of the large number of income classes. Yet even there it persists in diminished fashion.

62 Controlling health care costs

FIGURE 4

Sample plans illustrating income class-linked patient cost sharing

PLAN 'A'			
Income Categories	Deductible \$	Coinsurance Rate %	Catastrophic Limit % of Income
0-4,999	0	10	5
5,000-9,999	50	15	5
10,000-14,999	100	20	5
15,000+	150	25	5

PLAN 'B'			
Income Categories	Deductible \$	Coinsurance Rate %	Catastrophic Limit % of Income
0-4,999	0	10	5
5,000-9,999	50	20	7
10,000-14,999	100	30	7
15,000+	150	40	7

PLAN 'C'			
Income Categories	Deductible \$	Coinsurance Rate %	Catastrophic Limit % of Income
0-4,999	0	10	5
5,000-9,999	50	20	7
10,000-14,999	100	30	8
15,000+	150	40	9

SOURCE: Ontario Economic Council (1976, 45)

generate a charge, we have already observed that the impact of such charges on utilization appears to be small, if not, as in the case of hospital care, non-existent. It is true that design of supplementary private insurance may be more difficult to integrate with these plans, and that collections may be easier as part of the income tax system, but these advantages become insignificant if there is no deterrence.

At an operational level this type of plan involves information requirements not shared by the uniform-charge plans previously considered. To link health care utilization and income information one requires a common identifier as well as access to both sets of data. In view of the current political unease about confidentiality the only viable option would likely be to establish at the federal level a master list linking health care identifiers (such as OHIP

numbers in Ontario) to social insurance numbers. This type of linkage would facilitate the building of health care charges into the income tax collection process. But the necessity of linking two sets of data, each of which has been loudly criticized on grounds of inadequate protection of confidentiality,²⁴ suggests that administrative roadblocks could preclude consideration of any income- or income tax-linked scheme, regardless of other potential merits.

A second problem, one recognized by the Ontario Economic Council (1976), concerns the determination of incurred costs. Physicians' claims records are a ready source of medical service cost information which would require supplementation from patient claims records for patients receiving care from opted-out physicians. 'Free' drugs administered under plans such as British Columbia's pharmacare would also be relatively straightforward to monitor. The complexity of expenditure allocation is most pronounced when we come to estimate the cost of a day or an episode of hospital care, the dominant component of health expenditure. Hospital per-diems are unequivocally inadequate. Disease costing is prohibitively expensive for the myriad of diagnoses encountered in any given hospital.²⁵ Some tentative steps are being taken to improve upon crude per-diems while avoiding the complex detail and expense of disease costing (Barer, 1979; Barer and Manga, 1978). Hospital- and diagnosis-specific per-diems of the type being developed would greatly facilitate hospital cost allocation in income-linked schemes such as those described here; in their absence, allocation of hospital costs to particular patients will inevitably be arbitrary and inequitable.

These costing and information complexities aside, it turns out that the Council's plan (and other similar ones) come under familiar fire when set against our evaluative criteria. We established earlier that coinsurance and deductibles each embody an increase in patient risk-bearing (compared to a universal first-dollar plan), although many of the heavy users would likely be exempt from direct financial responsibility. This is no less true of the income-linked plans proposed by the Council; again, many or all of those who would have been exempted from direct charges under the other plans will find themselves exempt by reason of the income class into which they fit. For those not exempt and who do not purchase 'wraparound' private in-

24 This concern has most recently surfaced in Ontario through the activities of the Krever Commission on confidentiality of health records.

25 The limited attempts at disease costing include Babson (1973), Babson et al. (1975), Lay (1978), Evans and Robinson (1973), and Evans et al. (1975).

surance, the degree and extent of risk-bearing is illustrated in Figure 5 for a non-zero deductible income class and a representative Ontario Economic Council plan. It is clear that such a plan embodies a greater degree of risk-bearing than a standard deductible plan, but only through the range subject to the coinsurance. A Council-type plan requires a greater relative patient contribution than a standard coinsurance plan if total costs incurred are less than \$X in Figure 5. For costs greater than \$X, the Council's plan is a clear winner on risk-reduction grounds because of its built-in ceiling.²⁶ But in any guise the Council's plans involve greater patient risk-bearing than a universal first-dollar plan.

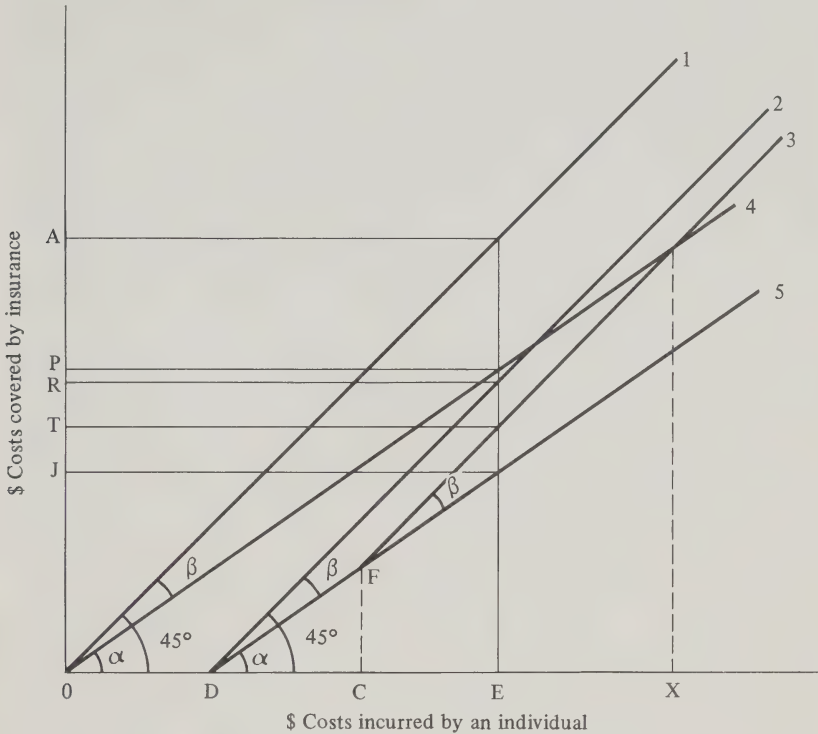
Consider our representative consumer with \$E in health care costs. Direct charges associated with the Council's plan would amount to \$AT. Assuming again that $E = \$300$, and letting $C = \$250$, the consumer would pay \$160.00, slightly less than under scheme 5 in Figure 5. While a ceiling is more commonly expressed in terms of the consumer's maximum financial liability, if we know the coinsurance rate and D, this can be converted directly to a level of incurred costs past which the consumer is fully insured.

Turning to effects on levels and patterns of utilization, the fact of the Council's plans being combinations of coinsurance and deductible charges suggests that utilization effects (or the lack thereof) will be similar to those described for each of these plans in isolation. Thus, direct exemptions by reason of income class, the purchase of private insurance, and countervailing provider behaviour will in combination again tend to mitigate any direct utilization effect. For example, it was noted earlier that if a deductible on hospital charges deterred admission, the hospital and attending physician retained considerable scope for adjusting lengths of stay of patients whose care expenses had already exceeded their deductibles. This type of manipulation would be a little more difficult under the income-linked scheme, since it would require that the hospital/physician be aware not only of the present and any past costs incurred but also of the patient's income class. But the widespread existence of discriminatory pricing before the spread of insurance in Canada and the United States suggests that practitioners are

26 Of course the actual placement of X, determined by the intersection of the coinsurance and Ontario Economic Council plans, will be dependent on income class, the coinsurance rate in the former plan, and the composition of the latter plan.

FIGURE 5

Consumer risk-bearing in Ontario Economic Council schemes



Note: Full first-dollar coverage (01); Deductible plan (0D2), deductible portion (0D); Ontario Economic Council plan (0DF3); Coinsurance (04); Deductible-plus-coinsurance plan (0D5), deductible portion (0D).

particularly adept at assessing patient wealth.²⁷ It is difficult to say, therefore, how pervasive this effect would be in preventing manipulation of lengths of stay. The same sort of reasoning could be applied to physicians' services. Of course once a consumer's ceiling is reached, that consumer is universally

27 In the United States, for example, studies have shown that physicians generally make an initial effort to assess the financial means of the patient and that this information influences their treatment decisions (Krizay and Wilson, 1974, 119-20: 'approximately 65% of the responding physicians try to determine, during the initial examination, what kind of health insurance coverage the patient has. More than 60% could recall deciding whether or not to hospitalize patients on the basis of their insurance

covered for the remainder of the period in question. Administratively, if the income tax return was employed as a collection and income class categorization vehicle, these schemes would be similar to third-party administered coinsurance collected annually, an option discussed earlier in this chapter. But the ceiling in these plans suggests that any direct utilization effects will be weaker (at least for some segments of the population) than under the annually collected coinsurance plan. Compounding that greater weakness is the nature of the linkage between use and charges. If, as we are supposing, the income tax return provides the medium for payment, those who engage a third party to complete their tax returns may be 'spared' the consumer-awareness role of direct charges. Of course one could equally argue that being aware of the user-charges linkage beforehand is a stronger deterrent than computing the magnitude of the damages afterwards.

Even if one believed that there were any residual influences of such an income-linked plan on overall levels of care utilization, it is clear that they would provide no more incentive to improved technical production efficiency than the uniform-charge plans discussed above. Since no individual provider can adjust reimbursed fee levels to attract new patients (even if such competition for 'market share' by price-cutting might be effective), there is no incentive or pressure on providers to seek out more efficient forms of production either aggressively, by expanding practice scope and income, or defensively, to keep from losing patients to more efficient providers.

This brings us to wealth transfer effects. Here, and here only, does the existence of income linkage allow some advantage to these plans. If we first make the simplifying assumption that the comparative (deductible and coinsurance) plans embody no exemptions, it is clear that making the deductible and coinsurance each a function of income class renders the Council's schemes more progressive than either. The schemes may also enhance the progressivity of a general revenue-financed universal plan, depending on whether the direct charge revenues substitute for a regressive tax such as

coverage.'). This suggestion that physicians modify utilization patterns in response to utilization coverage may appear inconsistent with the evidence above that health care utilization is insensitive to direct charges. One must, however, distinguish between individual and aggregate utilization, the latter being primarily dependent on the availability of hospital capacity and medical practitioners. Changes in relative insurance coverage appear to reallocate services from those without coverage to those with, while having minimal effects on total utilization — precisely the findings of Enterline et al. (1973) and Beck (1971, 1974). To infer aggregate response from individual responses is a methodological error known in elementary economics as the fallacy of composition.

premiums, or a progressive tax such as the income tax.²⁸ But even compared to deductible or coinsurance plans with exempt consumers, the Ontario Economic Council's schemes are superior on wealth transfer grounds. If we assume that those exempt from a deductible or coinsurance provision would also be partially or wholly exempt by reason of income class in the Council's schemes, the progressivity of the latter for non-exempt consumers is clearly a transfer in the socially optimal direction. The wealthy sick will subsidize care of the indigent sick (to the extent that they have not purchased private coverage) through their OHIP-related charges as well as their other tax contributions.

These wealth transfer advantages exist, however, only across income classes. While consistent with the public finance principle of vertical equity (equitable distribution of burden across income classes), the Council's proposal violates the equally important principle of horizontal equity (equal treatment within the income classes). One might interpret the distribution of health care burdens within an income class on the basis of illness as adding insult to injury.

It would seem, then, that with respect to the uniform-charge plans considered in this chapter the Council's type of scheme is most consistent with the social priorities set out earlier. Such a scheme has wealth transfer advantages over other uniform charge plans, and none of the plans affect any of the other three objectives in a socially desirable direction.

This discussion of income-linked direct charge schemes modelled on the Council's examples adopted the income tax return as the vehicle for both income class determination and payment/collection. In contrast, an income *tax*-linked plan would see the tax return retain only the latter function. There would be no explicit income class determination. Instead, calculation of financial responsibility for health care costs incurred would be subsumed within the larger income tax assessment. In particular, costs incurred during receipt of care in a given year would be equated with taxable health care benefits, and 'direct charges' for health care would then be determined by one's marginal tax bracket. Building in a ceiling on taxable health care benefits (similar to the present ceiling on Canada Pension Plan contributions)

28 While some low-income persons who are ill may end up paying more in charges related to health care utilization than healthy persons in upper-income classes, this pattern will not predominate. In general, the indigent sick will pay little, and at an aggregate income class level that flow will be overwhelmed by the reverse subsidization inherent in the charges levied on upper-income users.

would clearly make this type of plan more consistent in spirit with the Council's plans.

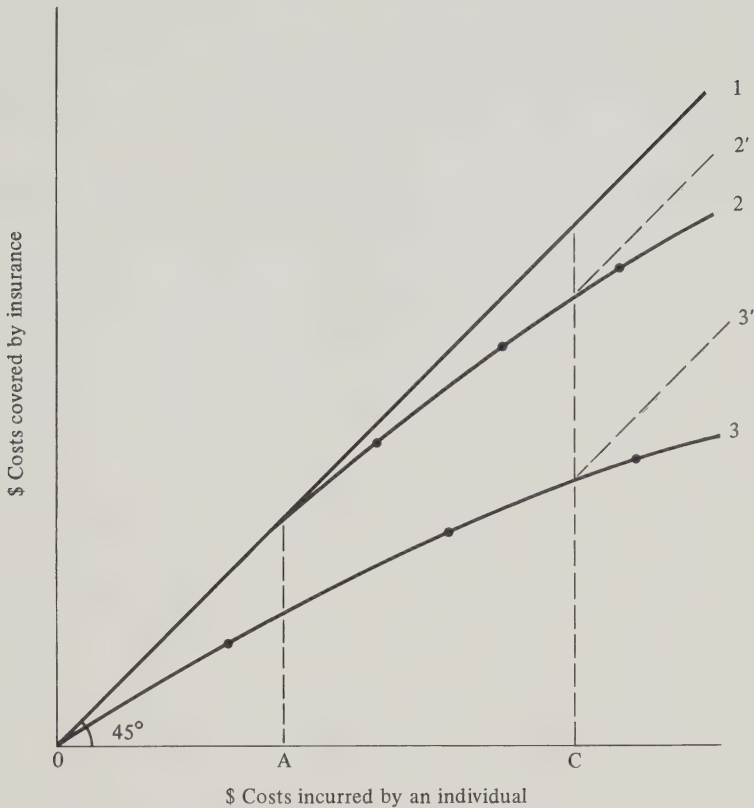
The major distinction, then, between the Council's type of income linkage and direct income tax linkage is that, in the former, incurred health costs lead directly to an increase in tax payable (for non-exempt consumers), while in the latter the increase is to taxable income and then only indirectly to tax payable. The conceptual distinction is not a large one, and while the burden for any given individual may differ between the two variants of income linkage (because, for example, with the income tax option one's income class could be influenced by a high level of health care expenditures), the two general schemes would probably have similar effects in terms of the four social objectives.²⁹

With respect to spreading risk, the computation of taxable income, including health care benefits, would now determine exemption status. For those not exempted, there would be a clear increase in risk-bearing compared to a universal first-dollar plan. Comparison with deductible or coinsurance plans is less straightforward, because any assessment depends on knowledge of a consumer's taxable income, the coinsurance rate or deductible level which is the object of comparison, and whether or not there is a ceiling on taxable health benefits. For example, if the income tax plan involved no ceiling, relative risks involved in that and the coinsurance plan would be directly determined by a comparison of the coinsurance rate and the consumer's marginal tax rate. The nature of risk-bearing for this type of plan is illustrated in Figure 6 for two different representative consumers. A consumer under plan 2 has no taxable income before health care benefits. OA represents the level of benefits which would give the consumer the maximum taxable income for which no income tax would be owed. Benefits (costs) accruing above that level would be subject, then, to 'direct charges' in the form of tax liability. The changes in slope represent changes in marginal tax rate as health benefits cause the consumer to move into higher marginal tax brackets. Plan 3 illustrates the situation for an individual who has taxable income before health benefits are added. For this consumer the changes in slope will be more marked than for the first consumer because the income tax

29 For example, with the straight council-type income-linkage scheme, a family in a low (say under \$5000) income class which incurs \$15,000 in medical and hospital expenditures would pay a maximum \$250 if the catastrophic limit is 5 per cent (see Figure 4). If, instead, the \$15,000 is added to income as a taxable benefit, the family could be liable for a significantly greater sum, depending on the nature of the deductions they are able to apply against gross 'income'.

FIGURE 6

Consumer risk-bearing in income-tax-linked schemes



Note: Full first-dollar coverage (01); for 02, 02', 03, 03', and 0A, see text.

program is progressive. Plans 2' and 3' are variants of plans 2 and 3 illustrating a ceiling past which no further tax liability accrues.

There is little to add about effects on levels and patterns of utilization. This plan will be similar to an annually collected coinsurance plan with a ceiling again dampening any direct utilization effects. The only possible distinction might be that building the benefits into the income tax calculations may further dilute the user charges/cost awareness link, because there would be no explicit 'health care costs payable' item in the tax payable

section of the return. Similar considerations to those mentioned for the income-linked schemes will also limit any significant impact on production efficiency. Finally, wealth transfer effects will be as described previously for those schemes with one very important exception. Under the income-linked plan, a low- or zero-income individual incurs nominal or no direct charges, while under the income tax-linked plan an individual with no other taxable income, but with very high health care utilization, would be raised to a higher tax bracket and expected to pay income taxes on these 'taxable benefits' accruing.

To this point we have considered income- and income tax-linked charges for health care services received (and costs thereby incurred). An alternative income-related direct charge might be built around the concept of 'health rebates'. In general, those who use no health care services during a given period would be eligible to claim rebates when filing their income tax returns. Partial rebates would accrue to those who used limited services.

Two possible means of linking such a rebate to income levels come to mind; we might call them the fixed rebate/variable cost plan and the variable rebate plan. In the former an income-variant share of incurred costs would be debited from a fixed rebate. Thus indigent persons might automatically receive the full amount of rebate regardless of medical service use intensity because of a zero 'coinsurance' rate. Low- and middle-income families might be charged on a sliding coinsurance scale of costs, while upper-income brackets would receive the rebate less fully charged costs incurred. The amount of the rebate places a ceiling on the amount of health expenditures charged back to any individual.

The variable rebate scheme would, as the name suggests, involve a rebate whose magnitude was inversely related to income class. Costs to be debited against that rebate could then take the above form (i.e. coinsurance rate varying according to income class), in which case the progressivity would be of a dual nature, or could be fully charged for all users. The former option contains the advantage that the indigent sick might still be eligible for some portion of the rebate. The income tax return would, as with income-linked charge schemes, serve as a payment/collection mechanism and as an income class categorizer. Thus, taxable income could be used to establish any given taxpayers's 'coinsurance rate'. This type of scheme is essentially a modified income-linked plan, with the rebate level for any income class serving as the ceiling on patient liability. Thus, it retains the essential characteristic of that class of plans: income is linked directly to utilization. This feature ensures that the plan is more progressive than coinsurance, deductibles, and per-service charges. It may thus improve vertical equity, though providing no

advantages in horizontal equity. Within any income class the healthy will gain at the expense of the ill. Wealth transfers relative to a universal first-dollar plan are difficult to assess without specific information about the source of tax revenue financing the rebate and about the incidence of illness across and within income classes. Furthermore, inherent in the variable rebate scheme is an added perverse twist: since the magnitude of maximum rebate is inversely related to income class, the degree of deterrence to which a patient is subject falls as the patient's income rises. Even if the coinsurance rate at which costs were to be debited from the income-linked rebate was linked and positively related to income, high-income users would reach a level of health expenditures above which they would no longer be subject to any deterrence far faster than would their lower-income counterparts.

With respect to the other policy objectives, the effect of a rebate will be similar to that of the income-linked plans described earlier. Since reductions in the rebate are essentially direct charges, individual risk-bearing will be greater than with a universal first-dollar plan. Rebates are not likely to have any significant impact on levels of utilization, for the reasons noted above. In addition, whereas non-income-linked direct charges would tend to discourage the poor from using needed services and encourage the rich (and/or their providers) to use (prescribe) marginal services, the progressive nature of the deterrence in the fixed rebate plan would tend to minimize even that utilization effect. Finally, effects on the efficiency of production will be unchanged from those described for all previous uniform-charge schemes.

In summary, then, the 'uniform' nature of the income- and income tax-linked charge plans and of the income-linked rebate schemes imposes a predictable (insignificant) impact on patterns and levels of utilization and on production efficiency. The rebate scheme will undoubtedly also raise total health care costs if the rebate is included in the public accounts as a health care expenditure item.

It is only on wealth transfer grounds that these income and income tax plans have something significant to offer. That is not to belittle this objective or their impact on it. On the contrary, we have seen that all the other uniform-charge plans not only were definitionally 'uniform' but were also uniformly devoid of any socially desirable impacts in terms of our four public policy objectives. The schemes described in this section suffer from most of the drawbacks inherent in the other more straightforward uniform-charge plans. In addition, they bring with them a number of administrative and operational problems of uncertain complexity. But they do contain the scope for improving upon the wealth transfer effects of coinsurance, deductible, or per-service charge schemes. Depending on the relationship between

income and the incidence of illness and on the progressivity of the sources of insurance plan revenue, such schemes would be superior in vertical equity to a universal plan financed, for example, entirely from premiums. All of them, however, would be inferior in both vertical and horizontal equity to a universal system funded more progressively.

It should by now be clear that direct charges uniform across providers (quadrants A and B of Figure 1) would have little socially desirable effect on risk-bearing, wealth distribution, utilization levels and patterns, or production efficiency. Patients will be no better informed about the efficacy of alternative treatments, and providers will continue to generate secondary feedback effects (which are likely to run counter to improved efficiency and lower costs) in reaction to any reduction in the numbers of patients beating paths to their doors.³⁰

Nothing short of wide-ranging policies directed at the very heart of the delivery system are required to affect utilization, efficiency, and thus costs. For direct charges to patients to moderate health care costs, such charges must somehow revise the currently perverse incentive structure in health care and promote improved delivery efficiency. Only if charges for comparable services vary across providers is it possible for a price competition/consumer choice/production efficiency/lowered costs chain to develop. The following chapters consider the use of differential direct charges as a cost control mechanism and as a means of facilitating achievement of our four social objectives without harming community health.

30 This reaction of providers to decreases in quantities demanded is not unique to health care markets. Some California residents have suggested that recent water rate increases represent an attempt by utility companies to make up for revenue lost due to successful state water conservation campaigns ('Water: penalty for saving', *Consumer Reports*, 43, No. 8, August 1978, 430).

4

Provider-determined differential charges

First doctor: What did you charge her?

Second doctor: £200

First doctor: What had she got?

Second doctor: £200

George Bernard Shaw,
The Doctor's Dilemma

Few propositions are more widely accepted in health care than the common observation that the existing financing structure provides no incentives, and in fact provides positive disincentives, for physicians and for hospitals to identify and weed out ineffective and unnecessary services and to improve the efficiency and lower the cost of their operations. Ultimately this derives from the obvious fact that health care costs are health provider incomes. As a result, any informed observer of health services delivery can draw up a long list of ways in which both efficiency and effectiveness can be improved, but can compile a list almost as long explaining why it is against the best interests, economic or otherwise, of providers to make such changes.¹

1 This is generally true for all providers of health services, whether the small practice owned and directed by a single physician, dentist or pharmacist, or the hospital or large clinic whose employees' salaries are paid from its total revenues. In our terms these employees — administrators, nurses, staff physicians, cleaning staff — are not

If any system of direct charges to patients is to succeed in revising the current system of perverse incentives for suppliers, patient decisions in response to charges must have direct effects on suppliers. Such effects will develop only if, as in the private marketplace, a provider's market share depends (inversely!) on his price compared to the prices of other providers. If charges to patients lead them to seek out lower-priced providers, the higher-priced will lose market, workload, and income. This will induce a search for more efficient, lower-cost forms of delivery and for lower health care costs overall. The market scope of the lower-cost provider will be expanded, that of the higher-cost provider reduced.

While this process may not take place in health care markets for a number of behavioural and institutional reasons and because of social policy objectives, two points must be emphasized. First, the entire process would hinge on price variation between suppliers. Consequently any schemes for charging patients will have to allow charges to vary across providers, unlike the proposals analysed in the preceding chapter. Second — and the basic message of this report — if this process does not exist, direct charges to patients cannot improve health system efficiency and cannot lower health costs. The only possible exception might be the very muted and dubiously useful deterrence effects noted above. Whether the mechanism of consumer-payment/price-competition/consumer-choice/provider-efficiency is in fact workable may be considered further; but if it does not work there is *no other* mechanism in the process to link consumer payment with provider efficiency.

For this mechanism to function, two sets of preconditions are necessary: market structure/conduct conditions and product-characteristic/consumer-behaviour conditions. The first set, encompassing provider organization and modes of behaviour, amounts to the display of price-competitive behaviour. If providers of insured health services (physicians, hospitals, pharmacists, dentists, etc.) are permitted to make direct charges to patients for a portion of the costs of their services, this charging behaviour must be truly independent, open, and advertised. For example, a provider association that promulgated an 'accepted' direct charge to patients and discouraged its members from undercutting or advertising alternative charges would of course destroy the effect of the proposal. Under these circumstances, the direct

themselves providers; the hospital or the clinic is the provider. Strictly speaking the physician or the pharmacist is not a provider either, it is the practice (including physician, aides, and facilities) or the pharmacy which provides services. But in these cases the professional individual plays such a large role in the functioning of the institution that the distinction is frequently blurred.

charge to patients would become simply a device for increasing the price of services without imposing any competitive pressures on individual providers. Of course such formal collusion would presumably be a criminal offence under the most recent revisions to Canadian anti-combines legislation (Canada, 1976). But a wide variety of techniques for informal collusion and 'conscious parallelism' persist (see, for instance, Moore, 1973; Stanbury and Reschenthaler, 1976). How the real social interest in close co-operation between health care providers for purposes of continuity and co-ordination of patient care, consultation over therapy, and continuing education can be reconciled with the social interest in complete separation of economic behaviour is not clear. It is difficult to believe that physicians, for example, coming from a background of many years of training in close proximity to ensure a common value system, subsequently working together every day in clinics and hospitals performing similar tasks, and further extending their contacts socially, can avoid co-operating on prices when it is so obviously in their interest to do so. Yet the usefulness of direct patient charges as an inducement to provider efficiency is critically dependent on the absence of collusion.

The prospects for energetic price competition between hospitals seem even more limited, in view of the small number of hospitals in many market areas, the close and socially encouraged co-ordination of hospitals in the same region for purposes of patient care, and the close linkage through hospital unions and medical staffs.

The absence of provider collusion, whether explicit or implicit, is a very strong necessary condition. But it is not sufficient to ensure that direct charges will affect efficiency. The patient must also be sufficiently informed to be able to judge the relative merits of the alternative service packages offered at different prices. Alternatively, regulatory mechanisms must exist to ensure adequate quality and safety standards for the services provided. If the patient cannot judge service quality, and instead resorts to a rough judgment of quality by price, the whole mechanism breaks down. If, on the other hand, consumers are quite price-sensitive, but producers find it profitable to provide inferior quality services without fear of detection, the private and social interest in health standards is threatened. Low-quality health care is not merely less of a 'good' than high-quality care; it can often be a positive 'bad', that is, worse than none at all. But patients cannot in general be relied upon to distinguish good from bad.

Very likely, therefore, the party responsible for preventing collusion and monitoring quality would play a critical role in the success of a differential charge system. In this chapter we consider schemes for differential charges as

much like the market process outlined above as possible, with a view to identifying how all preconditions might reasonably be satisfied and how informed consumer choices might lead to effective and efficient levels of utilization. As indicated in Figure 7, three routes by which provider-determined differential charges might be introduced are major-risk medical plans, extra-billing systems, and a redefinition or 'repackaging' of health services and institutions.

MAJOR-RISK MEDICAL INSURANCE

Major-risk medical (MRM) insurance² has traditionally referred to a deductible plan in which a limit is placed on out-of-pocket expenses. This concept differs in two significant ways from the uniform charge deductible discussed earlier. First, it has generally been associated with differential charges, MRM policies generally being sold by private insurance companies exercising no influence over provider charges. The private company reimburses the patient for all or a portion of expenses above the limit after these expenses have been paid by the patient. The uniform charge deductible discussed above envisaged the provider being reimbursed by the plan at uniform rates and the patient later reimbursing the plan. It is doubtful if uniform charges can be maintained in any system in which the program reimburses patients rather than providers.

Second, in MRM insurance the deductible is relatively large (well above the annual health care expenditure of the average individual), so as to protect only against catastrophic losses resulting from the treatment of serious

2 This should not be confused with 'maximum dollar limit' coverage (Krizay and Wilson, 1974, 19-22), which embodies an opposite principle – leaving the patient at risk for payment of extremely high costs. The plan entails a limit to the extent of coverage, e.g. \$10,000 or \$50,000, in any period. Depending on that limit, of course, very few consumers will end up in the 'over limit' category. But whoever ends up in that category is truly left to his own devices. The degree of risk-bearing on the part of the consumer in this type of plan is obviously greater than that for a universal plan. Wealth transfer effects are as for all direct charge plans – from the ill to taxpayers. No obvious incentives to efficient production come to mind. It is for utilization effects that this plan is most interesting, because even the potential first-line deterrence will most likely be absent. Any patient who exceeds the coverage limit is likely to be so ill as to have little decision-making scope with regard to receiving or avoiding treatment. And one would be hard-pressed to argue that the marginally necessary consumption of care would be deterred. This plan seems, then, to hold some appeal only for those responsible for providing the insurance coverage.

FIGURE 7

Methods of increasing patient participation in payment for health care:
provider-determined differential charges

		Charges determined by	
		Non-provider	Provider
Level of charges across providers	Uniform	Coinsurance Deductibles Per-service charges Income- and income tax-linked proposals	Coinsurance Per-service charges
	Differential	D	Major risk medical Extra-billing Service repackaging

illness.³ The intent is to insure against large losses but to require patient payment of a high (in this case, 100 per cent) proportion of costs incurred up to the amount of the deductible. In effect, then, with MRM insurance the provider is free to charge the patient whatever fee he deems appropriate, but the patient is eligible for reimbursement of all expenses above the limit for a given time period. The provider determination of fees and the lack of a

3 What constitutes 'large' expenditures varies over a wide range. A recent US government report on catastrophic illness (United States, 1978) defines a catastrophic illness as one generating expenses of \$5000 or more a year and estimates that persons with such expenses (1.2 per cent of the population) generated \$22 billion of health expenditures in 1974, or 19 per cent of total US health expenditure. If the catastrophic cutoff is lowered to \$3000, it includes 2.2 per cent of the population and involves a cost of \$31 billion, or 27 per cent of US health cost. Presumably the high-expenditure definition underlines some US thinking about the aptly named 'catastrophic' national health insurance proposals. By contrast, the Ontario Economic Council's deductible proposal, mentioned above, runs from zero to \$150 depending on income class, and the deductible levels used by Beck and Horne (1976) in partitioning the Saskatchewan medical care claims data are in the \$25 to \$100 range.

uniform fee schedule place this scheme in quadrant C of our original matrix (Figure 1).

While the deductible limit has been the usual form of MRM insurance, other options embodying the same principle might be suggested. Retaining the deductible for illustrative purposes, we might wish to distinguish a fixed-limit scheme from an income class variant scheme in which the size of the deductible is a function of income. The major difference between the two would be in the wealth transfer effects. Within either scheme, however, any number of alternatives might be employed. As long as a ceiling exists above which patients are not liable for additional expenses, one may require them to pay all or a fixed proportion of expenses up to the ceiling, or a proportion varying with expenditure level. The income-linked major-risk options obviously resemble the Ontario Economic Council's scheme discussed in the previous chapter. While the Council's examples involved relatively low deductibles, their general framework is analogous to the spectrum of major-risk options suggested above. The major operational difference is that the Council's scheme was offered in a setting of negotiated hospital budgets and uniform fee schedules which provide some control over total costs.

Compared to universal first-dollar coverage of health costs of the present type, the MRM approach is obviously inferior on our first two objectives. The insured population is subject to risk, particularly if the level of expenditure at which reimbursement begins is set at several thousand dollars. If the MRM plan has a percentage coinsurance feature, of course, the risk is even greater. And compared to first-dollar coverage, MRM insurance transfers wealth from the ill to the well within income classes. It may or may not redistribute wealth across income classes, depending on whether or not the deductible limit varies with income. Such variation presumably requires some integration with the tax system; and this linkage would be more difficult to achieve with a private insurance-based MRM system. Insofar as health (medical and hospital) expenditures are higher for lower-income groups, then, the private MRM approach also transfers wealth from low- to high-income groups.

Against these weaknesses, MRM advocates in the United States claim advantages on the other two policy objectives — service volume/mix and technical efficiency. The existence of the deductible is supposed to discourage patients' use of 'frivolous' and unnecessary care, and the price sensitivity of consumers paying 100 per cent of their bills up to the limit is supposed to encourage price-shopping among providers who would now be allowed to set charges independently. This in turn is supposed to induce provider price competition over market share, lower service prices, and enhanced technical efficiency. Thus, there are two issues to be explored in

comparing MRM plans with first-dollar coverage. Do they influence total utilization? And do they influence the distribution of utilization, or market share, among providers?

On the first issue, the sensitivity of overall utilization of health care to the existence of a deductible or an MRM-type plan, all the considerations discussed earlier in the context of coinsurance apply. Many patients (including the highest users of care) will have to be provided full coverage due to age and/or poverty. If the deductible limit is large, private purchases of fill-in coverage will result. And if utilization actually does fall, providers can respond by generating increased utilization and/or raising charges for their remaining patient load. Patients whose use exceeds the deductible, or who are excused payment, will not likely be price-sensitive. The main difference compared to coinsurance will presumably be that the deterrent to initial contact with the health system in any year is greater because the patient bears all the initial costs. But once the deductible is exceeded there is no further deterrence. Thus, in the sample coinsurance computations one would exclude a higher proportion of total costs from the deterrent effect — expenditures generated by the excused population (poor/aged) and those generated by the patients who had exceeded their deductible — but would apply a larger deterrent effect to the remainder. Unless the deductible limit is large, \$1000 or more, almost all hospitalized patients will exceed the deductible, so that all their medical services will be insured as well. Thus the deductible deterrent will apply to the health services consumed by the non-aged, non-poor, and non-hospitalized population — not an impressive amount of the total.

For this group, MRM-type plans may indeed affect utilization. Newhouse et al. (1978) estimated that a deductible of \$200 would lower the 'medical care' expenditures for the population they studied by 15 per cent compared to a \$50 deductible, and a \$500 deductible would reduce expenditure by a further 13 per cent. Another 7 per cent reduction occurs with a \$1000 deductible. No data are available on use at full first-dollar coverage. On the surface, this appears to suggest that low deductibles have substantial potential for limiting use. The issue of necessary or unnecessary use is of course not addressed.

Regrettably, the argument sags when one places any weight on it. The study population, though large, is all employed and does not include dependents. Thus the aged, children, and the chronically ill are all excluded. Further, pregnancy-related expenses are excluded, and neuropsychiatric care is included only at 50 per cent. Moreover, data on out-of-pocket expenditures by those of the insured population who did not exceed the deductible were not available and had to be estimated.

In addition to these exclusions from total expenditures, there are well-known biases when the data are drawn from commercially insured populations. First, insofar as purchasers of insurance have any knowledge of their general health status, people who perceive themselves to be in poorer health will choose more complete coverage. This self-selection bias, as the authors note, will lead to observed higher expenditures per person for the lower deductible groups whether or not overall use has changed. Provider responses, if any, have the same effect; if the high deductible reduces use among part of the population and providers generate new utilization across the whole group, we shall observe a shift in the *distribution* of services through the population but not a change in totals. Finally, as Roemer et al. (1972) showed in a study of alternative insurance plans in the United States, private for-profit insurance plans behave as profit-maximization dictates and become skilled at selecting healthier individuals or groups as customers. How the screening is done is not always clear, but the effects show up clearly enough.

Thus Newhouse et al. showed only that healthy people may moderate their use of health care in response to deductibles; but due to certain biases we cannot be sure. What is established, however, is that the population studied is not representative of the average user of health care. Estimated annual 'medical' expenses (which appear to have been medical and hospital, though hospitals are not mentioned) ran between \$115 and \$180 for the study group (in 1975 dollars) compared with the \$601.16 average per capita for the US population as a whole (derived from United States, 1977, 343, average of fiscal years 1975 and 1976). The persistent focus of US 'demand' studies on employee groups is understandable in view of the nature of US data; what is less understandable is the nonchalance with which such estimates, warts and all, are generalized to the whole population. Even if one had total confidence in the validity of such estimates for the group studied and were able to ignore provider behaviour and supplementary insurance influences on utilization, generalizing the behaviour of healthy employee groups to cover the 0.6 per cent of the US population that is institutionalized and incurred catastrophic (over \$5000 a year) expenses (United States, 1978, vi) is difficult to justify. That group accounted for 9 per cent of total US health expenses; as noted above, 27 per cent of all such expenses went to the 2.2 per cent of the population that generated more than \$3000 each in costs in 1974. These people would appear to be quite sick. It is doubtful if their response to a \$200 deductible would parallel that of a commercially insured employee group.

There does not, therefore, appear to be any evidence for the view that an MRM insurance plan would significantly limit unnecessary use more than

would a first-dollar full coverage insurance system. This does not in itself rule out the possibility that fully self-paying consumers could induce efficiency-enhancing competitive behaviour among providers. That is presumably the possibility in Feldstein's mind when he suggests (1976) that one of the major objectives of a public insurance program ought to be the prevention of catastrophic losses and then goes on (1977) to suggest MRM insurance as a viable means of tackling the 'high cost of hospitals'. The implication is that if public insurance focuses on risk-spreading, the private market-place can look after efficiency and price-determination. This assumes that the only current impediment to the smooth functioning of competitive markets in health care is the existence of 'excess' insurance (resulting, as noted above, from the tax structure). It would be hard to think of a less realistic assumption.

An MRM plan would indeed facilitate avoidance of a large public tax burden, but since total costs are not likely to be significantly affected (at least not significantly reduced) that implies no more than a shifting of burdens. One cannot expect both to control costs and to ensure that 'the current freedom of physicians and hospitals would be preserved' (Feldstein, 1977, 54). The differential nature of the charges implies an open-ended fee schedule. While there may be some deterrence from the direct financial risk shifted to the potential patient, the least useful services may not be those deterred. And while it is true that 'no family would be deprived of care because of inability to pay' (ibid, 52) (assuming an appropriately designed MRM plan), it does not follow that there would not be a tendency for some families to avoid seeking necessary care. It is certainly true, as stressed above, that there is ample evidence of the provision of marginally useful or wholly inefficacious health services. But to use this as evidence that deterrent-induced reductions in utilization (should they occur) will not affect health status (as is suggested, for example, by Newhouse, 1976) it must be assumed that the inefficacious services will be the ones deterred. This in turn entails that patients know which services these are.

In addition, the secondary feedback of the provider who reacts to a slowdown in patients by increasing services per patient or per episode now has a further dimension. The potential for price discrimination — charging the remaining patients more per service, depending on the provider's perception of their ability to pay — works against the initial deterrence as well. Thus, there is no reason to believe costs will fall or even that they will not rise. A fall in costs still requires a fall in provider incomes, *ceteris paribus*, and the differential-charge provision allows greater price freedom than the present universal plan with a negotiated fee schedule. Finally, a high limit will minimize the number of consumers covered universally in any period. But the

82 Controlling health care costs

feedback routes noted here suggest that an MRM plan will result in more patients reaching a given ceiling than would have been the case under a full first-dollar plan.

In fact, reflection on Canadian and US experience since 1971 suggests very strongly that an MRM insurance program along the lines recommended in the United States would have led to significantly greater cost increases in Canada than actually occurred. As noted in chapter 1, Canadian and US health costs moved in parallel through the two decades prior to 1971, steadily increasing their share of national income. Since 1971 this share has stabilized in Canada but has continued to rise steadily in the United States. The Canadian experience appears to be the result of centralized provincial bargaining of physician fees and direct provincial control over hospital budgets. The evidence seems quite conclusive that these mechanisms have controlled costs and the market forces in the United States have not. To argue then, as do Newhouse et al. (1974), that a US national health insurance system with full first-dollar coverage would in fact *increase* costs in that country (to, as suggested, 11 per cent of GNP) suggests either total ignorance of the Canadian experience or a judgment that the United States would be significantly less competent in running such a program.

The most probable result, then, is that an MRM insurance program, of the type recommended in the United States would lead to significant increases in the cost of health care, and probably to reductions in efficiency as hospitals, relieved of direct budget control, set their own charges. The plausibility of the 'free market' hypothesis that these forces could be offset by price-consciousness of buyers of care and price competition between physicians and/or hospitals seems quite at odds with everything known about the structure and behaviour of the health industry. The MRM approach thus offers cost increases and probable deterioration of technical efficiency, little or no gain on mix and level of utilization, increased consumer risk, and perverse wealth transfer. That is not an impressive policy.

EXTRA-BILLING BY PHYSICIANS

Physicians' services are an area where a number of proposals for differential direct charges to patients have been brought forward, particularly by medical associations.⁴ In general, and under reasonable assumptions, these proposals

4 See for example, 'Doctors want patients to pay them directly', and 'Ontario MD's Fight "OHIP conscription"', in *Hamilton Spectator*, 23 June 1977, 2, and 9 May 1978 respectively; see also Dr Donald Aitken's testimony before the Select Committee of the Ontario Legislature on Health Care Financing and Costs, 1 August 1978. Similar

can be shown to raise expenses for physicians' services. On reflection this should not be surprising, since expenditures on physicians' services equal physicians' gross incomes. Such proposals are apparently intended to avert one-to-one bargaining between physicians and governments in fee and income setting, in which governments have turned out to be surprisingly tough bargainers, and to afford physicians the scope for increasing medical prices and incomes at the expense of patients, who are, individually, much weaker bargainers.

The discussion in the public arena, like that in this section, has focused exclusively on extra-billing by physicians. But it can be applied to providers of other health services with appropriate modification. For example, under private insurance plans dentists already have the power to impose differential direct charges on insured patients. Each dentist sets his own fees and bills the patient, who is then reimbursed by the insurance company. Dentists can and do charge above this reimbursement rate, and at their discretion can charge patients different amounts. Hospitals as providers would be in a more complex position. But a system of hospital reimbursement that allowed each hospital to introduce its own schedules of direct charges to patients and did not treat the resulting extra revenues as offsets to the provincially provided budget would have similar effects. Of course next year's provincially provided budget would have to be independent of this year's actual costs, which would presumably increase to absorb the revenues from extra direct charges. Insured pharmacy services fit simply into this discussion; discretionary extra-billing would occur if the insuring agency reimbursed drug costs plus a flat dispensing fee and individual pharmacists set their own fees at or above this level. The analysis of extra-billing by physicians in this section thus applies, with some modifications for hospitals, to each of the other classes of providers. The balance between the different dimensions of quality may, however, vary among providers of different types of health services. Moreover the potential for manipulation of 'practice style' so as to increase patient throughput is probably less for hospitals and pharmacists than for physicians and dentists. Even in these settings, however, hospital managements can influence lengths of stay and patterns of service use, while pharmacists can establish record systems designed to draw patients back to the store. The differences appear to be ones of degree rather than kind.

views emerged from the fall 1977 meetings of the Ontario Medical Association, and have been a consistent theme of *Medical Post* editorials (e.g. 'Making the user pay', 13 September 1977, 12).

The most extreme such proposal would allow individual physicians to impose, at their own discretion, extra charges on some patients (over and above the negotiated reimbursement schedule) but not on others. This would involve a return to the pre-insurance pattern of price discrimination, which, as pointed out in Kessel's (1958) classic paper, is a powerful device for increasing provider incomes without a loss in volume of utilization. In general, it permits the physicians to charge more to those whose utilization is least likely to be affected thereby ('those who can afford it') and less or nothing (additional to the schedule) to those most likely to respond. Such a policy can easily be justified on medical or humanitarian grounds as ensuring that no one goes without needed services because of inability to pay and the burden of health care costs is loaded onto those perceived by physicians as most able to pay. The economic effects are clear, however. Such a policy permits individual physicians to achieve the maximum possible increase in the average price of their services consistent with little or no reduction in demand for their services.⁵ Since price rises and utilization stays constant, medical costs and medical incomes rise. Moreover, risks of price competition between physicians are minimized if each physician can claim to have no uniform rate of extra-billing. Hence information on the effective prices of any given physician's services will be very difficult to collect or transmit; for an individual patient, comparative price shopping will be virtually impossible.

There is a direct link between this type of proposal and the present group of opted-out physicians in Ontario. In fact, the proposal is quite simply a case

- 5 In the jargon of the economist, the policy permits physicians to absorb any consumers' surplus from the patient-initiated segment of utilization. Extra-billing for discretionary services is an add-on resulting from the agency relationship between patient and supplier. Of course the ability of physicians to capture that consumers' surplus may be somewhat restricted in an environment such as that currently existing in Ontario, where only a segment (currently around 19 per cent) of the total practitioner population has opted out and can extra-bill. Wolfson and Tuohy (1979) found, in fact, that the patient loads of a survey sample of opted-out physicians in certain specialties were significantly lower than those of their opted-in counterparts. If only opted-out physicians are permitted to extra-bill, and if those physicians are few in number, so that patients have an equally accessible, no-out-of-pocket-cost alternative, we might expect some diminution in the extra-billing impact on medical costs. However, Wolfson and Tuohy also found considerable variance between specialties in the ability of opted-out physicians to compensate for reduced patient loads. Opted-out general practitioners and medical specialists, in particular, seemed to be losers as a group in terms of net income levels, while the mean net income of opted-out surgical specialists and psychiatrists was higher (but not significantly so) than that for the comparable opted-in practitioners.

of universal opting-out, except that the 'opted-out' label would become excess baggage, because if everyone opted out of a plan it would be defunct.

The opting-out system for Ontario physicians involves differential charges determined through patient-provider negotiation⁶ and paid by the patient. OHIP subsequently reimburses the patient in an amount equal to 90 per cent of the Ontario Medical Association fee.⁷ Each physician must choose between opting in (and accepting 90 per cent of the fee schedule as payment in full) and opting out.⁸ Thus the Ontario patient (at least at present) retains the right to choose to accept point-of-service charges (if he can locate an opted-in physician). If patients were generally aware of this right and of the existence of price differentials between opted-out physicians, this situation could impart a degree of competition for market share. In fact, there seem to be differences in practice patterns between opted-in and opted-out physicians (Wolfson and Tuohy, 1979), but they are largely the result of the differences between physicians in their scope for generating feedback effects.

The increase in medical costs resulting from a policy of partial or whole-sale extra-billing arises because currently negotiated fee schedules are a floor upon which physician charges are additive; no physician is likely to pay rebates to his patients! One might imagine, however, a regulatory scheme whereby the floor could be adjusted downwards (or not raised as quickly over time) until the sum of public plus private payments reattained the level that would have been achieved by the public system alone in the absence of extra-billing. The only difference would be that while the burden of public payments is distributed through society by the tax system, that of private payments would fall only on the ill and would be proportionate to physicians' perceptions of their ability to pay (as qualified by social relationships and other factors). But such a system would be unstable, because any increase in

6 'Negotiation' may at times be a misnomer, however. The patient may be informed of the provider's rate of extra-billing and left the choice of accepting it or going elsewhere.

7 The result of recent negotiations between OHIP and the OMA has been the apparent abandoning of the relationship between the OHIP schedule of benefits and the OMA fee schedule (see, for example, *Globe and Mail*, 14 April 1978). OHIP has rejected the latter, which involved a 36 per cent increase in the fee schedule (*Globe and Mail*, 31 January 1978), in favour of an average 6.25 per cent increase in benefits. Thus, while OHIP is still willing to reimburse physicians at the rate of 90 per cent of the *benefit* schedule, the correspondence to the fee schedule will no longer be as straightforward.

8 For a discussion of certain exceptions and a more detailed consideration of opting out, the interested reader is referred to Wolfson and Tuohy (1979).

general fee levels by physicians above what the government would otherwise have negotiated would have to be met by a cut in the state share. The result of such a balancing policy would be a fairly rapid diminution in public payments, followed by a resurgence of private insurance, and finally a full-scale private system with no control at all over medical costs — precisely the situation that the Hall Commission studied and rejected and with which US policymakers wrestle so hopelessly today. It seems fairly safe, then, to suggest that universal extra-billing will do nothing to enhance cost control or to lower utilization levels. Furthermore, this type of open-ended system is perhaps the least likely candidate yet considered as a catalyst for efficiency improvement. Consumer risk-bearing will be greater than with a universal and first-dollar plan, and, as noted above, there will be a transfer of costs from taxpayers to the ill. To the extent that physicians are adept at assessing income class, it may at least be a progressive transfer. On a first pass, however, extra-billing has little to offer anyone but the providers.

A more interesting line of argument in support of extra-billing by physicians starts from the proposition that physicians' services are not homogeneous goods. Some physicians are better than others and should be paid more. Alternatively some physicians spend more time with patients and provide a more humane sort of service. Again, patients should be able to pay for this extra service if they so desire. The argument is a complex and sophisticated one and has several strands that need to be disentangled before it can be evaluated.

First, one must describe the differences between medical practitioners upon which differential reimbursement can be based. There seem to be three dimensions to this, which we may label quality, ambience, and style. Quality is an outcome measure: one surgeon may be inherently of higher quality than a second, for example, if patients are likely to experience better outcomes (lower mortality rate, post-operative infection rate, recovery time, degree of subsequent disability) if treated by the first (independent of time or effort intensity of treatment). Ambience on the other hand assumes equivalence in the ability of different practitioners to effect a given therapeutic outcome but suggests that some practitioners may spend more therapeutically irrelevant time with the patient, be more polite or supportive, invest in more pleasant surroundings, and/or generally provide a wider range of services of value to consumers apart from their therapeutic impact. It is, in effect, a dimension relating to the treatment of the patient as a consumer, to be distinguished from style, which relates to the treatment of a patient's condition. Style concerns the pattern of treatment during an illness episode rather than the outcome of the treatment or the amount of time per contact. An

episode may be resolved by a physician over only one or over several relatively long visits, with a careful diagnosis and treatment plan; alternatively the care may consist of a larger number of short visits with a high rate of referral for diagnostic or specialist services. The latter style is strongly favoured by the current structure of the fee schedule, although the former may be less costly overall (particularly if patient waiting time is taken into account) and may be better medicine in the sense of achieving a therapeutic result with less intervention. But a policy of restraining medical costs by holding down fee schedules generally penalizes physicians who choose the long-visit style, and relative income pressures will tend to force most practitioners towards the 'revolving door' style.

Each of these types of differences between physicians represents a genuine issue for physician reimbursement policy, but their linkage to direct patient charges is much less clear. Taking the first dimension, quality, in the sense of ability to achieve a therapeutic outcome, differences undoubtedly exist between physicians and are measurable with more or less difficulty depending on the specialty. But it does not follow from this that direct patient charges should correspond to these differentials. In the first place a system of permitting individual physicians to evaluate their own quality level and assess charges accordingly is patently absurd. Since patients cannot in general assess the relative quality of practitioners (in the outcome sense) — if they could there would be no need of licensure to protect the public against incompetence — there is no ground for relying on patient selection to generate a structure of relative charges correlated with quality. Thus a system which allows physicians to determine their own charges on 'quality' grounds reduces to the previous proposal of independent practitioner control over direct charges, but with a different cover story. Every physician will designate himself as more or less above average. We would not expect to find physicians declaring themselves below average, despite the mathematical fact that about half the profession must be so, and offering their patients rebates from the official fee schedule for that reason.

One might, however, make a case that quality differentials are real and ought to be rewarded to encourage general improvement. Merit bonuses to physicians identified by their peers as superior seem to be justifiable. But this would require setting up a peer review/monitoring committee not unlike those operating in hospitals. This in turn would require finding a group of physicians willing to sit in judgment on the practice behaviour and quality of their peers. The general reticence of members of the profession to testify against one another in malpractice suits suggests that construction of such a committee would not be easy. In addition, it would be subject to the in-

fluence of mutual backscratching if committee membership changed every three or five years.

Even if satisfactory answers were found for all these problems, it is not at all clear that prices to consumers should be used to ration access to superior physicians. That implies a deliberate endorsement of the notion of better care for the wealthier. Presumably what one really wants to do is allocate superior physicians to more complex cases, but that has nothing to do with patient charges, unless one believes that the illnesses of higher-income people are always more complex! Moreover, a merit bonus system on the current mode of reimbursement could skew reimbursement towards higher-quality physicians without raising the average. One could negotiate a base rate of reimbursement plus a merit rate such that the total package remained within social expenditure targets. Charges to patients, on the other hand, are additive to the public reimbursement level as a floor and thus inevitably raise total medical costs (unless of course the public floor is lowered each time private charges are raised – the unstable case already analysed). Thus the existence of quality differentials may be a justification for a peer-determined merit award system but provides no support for patient charges because there are no grounds for assuming that such charges would in fact correlate with quality, and because one would not wish to ration access to quality in that way even if they did.

The issues of practice style and of the biases intrinsic to fee schedules in general and to tightly negotiated ones in particular suggest a serious flaw in the concept of fee-for-service reimbursement. Fee schedules generally penalize the 'few long visits' mode of treatment compared to the 'revolving door' mode. Any argument for direct patient charges as a corrective, however, implicitly concedes that physicians have income targets and if these targets are not attained by increased negotiated fees, and direct charges are not permitted, there will be shifts in billing patterns to increase physician gross incomes. The resulting shifts in billing patterns may not be sufficient to reach target incomes either; they may result in more costly forms of care (unnecessary referrals or diagnostic tests) and lead to physician dissatisfaction with 'revolving door' medicine.

This scenario may well be an accurate description of medical practice in Ontario and elsewhere. But it is not so obvious that direct patient charges represent anything more than an argument for higher fees from whatever source. The implication is that if negotiated fees were raised, physicians could attain their income targets and slow down the revolving door with gains to themselves and to society. Since government in its presumed unwisdom refuses to raise these fees, direct patient charges are the only alternative

means for achieving a fee schedule increase. One can therefore evaluate the argument for higher fees independently of the argument for source of payment.

There is certainly a considerable amount of evidence to support this general description of the relation between fee levels and physician-induced utilization, particularly the close negative correlation across Canadian provinces between average annual fee schedule increases since 1971 and utilization or price-deflated billings per physician (Evans and Wolfson, 1978, 17). It is less clear, however, that in the short run a fee increase would cause the revolving door to slow down. Physicians in Canada now appear to feel 'poor' in relation to their historical peak relative incomes of 1971, since billings have not kept pace with inflation and faster rates of patient turnover have not been sufficient to maintain target incomes. Thus a fee increase, whether paid by government or by patient, would very likely be absorbed as a step toward restoration of 1971 relative income levels without any corresponding shift back to a style of less intensive utilization. But suppose this were not the case and in fact higher fees would moderate patient turnover rates. Then the argument for fee increases amounts to saying that any attempt to negotiate fees below what physicians collectively find satisfactory will lead to inappropriate utilization/billing patterns unless offset by charges to patients. But under fee-for-service reimbursement what alternatives (to holding down negotiated fees) exist for constraining medical costs? At root, the practice-style direct charges argument is a rejection of any constraint on physician fees/incomes, an argument for a professional blank cheque drawn on the patient as well as the government and backed up by the threat of procedural manipulation. The argument has merit, but in fact it also becomes a powerful argument for modification or abandonment of fee-for-service reimbursement. If control of medical expenses cannot be achieved under fee-for-service without serious perverse effects on practice style, then either one gives up control (by letting physicians set their own fees and collect from patients the share that government will not pay) or one gives up fee-for-service. Asking patients to foot the bill for avoiding the revolving door appears to amount to bribing the physician to avoid procedural overservicing for a given episode and to reduce patient inconvenience.

Extra-billing for style clearly fails to help achieve any of the four social objectives. In particular, the necessary preconditions for direct charges to function as an incentive to efficiency are not satisfied, because patients will not generally be able to judge the effect of alternative styles on a given illness.

Only on the third dimension, ambience, can it reasonably be argued that

the consumer information precondition would be satisfied, since by definition ambience has no therapeutic value. Patients generally will be able to compare different types of ambience and pay accordingly. But the link to production efficiency requires the provision of variable-ambience, equal-therapeutic-outcome alternatives: this question takes us in the next chapter into non-provider determined differential charges. If the issue were simply one of extra-billing for ambience, one could appeal to consumer sovereignty and let the market clear. But we are judging direct charges according to four criteria. Only if the existence of ambience alternatives is ensured by a third party is there any chance that cost/utilization containment and efficiency can be improved. For that reason the topic is considered in greater detail in the following chapter.

Extra-billing on whatever justification (with the possible exception of ambience) seems to hold little appeal as a tenable instrument of cost control, promotion of efficiency, risk reduction, or socially acceptable wealth transfer.

SERVICE REPACKAGING

If one of the most popular and apparently straightforward avenues of differential charges — extra-billing systems — is likely to hinder realization of our policy goals, one may reasonably ask if *any* proposals for direct charges to patients could satisfy the preconditions necessary for a competitive market process. While there appear to be a limited number of such proposals, they all require ‘repackaging’ of the health services involved, and most appear to require new definitions of the delivery organization as well.

Probably the clearest example of a differential direct charge satisfying the product-characteristic/consumer-information condition is the prescription drug insurance employed in Saskatchewan and discussed in Evans and Williamson (1978). This program distinguishes between the process of drug dispensing and the provision of particular drug ingredients. Since patients are not in general competent to evaluate either their own need for a drug or the relative merits of one drug over another (otherwise prescriptions by physicians would not be required for drug purchase), the insurance plan covers ingredient costs in full without any charge to the patient. The dispensing service on the other hand is a relatively uniform activity which consumers may quite reasonably be expected to choose to receive from one pharmacy or another on the basis of price/cost information. Certain basic quality standards must be assured by regulation, but these do not differ in kind from those required in retailing generally, such as protection against short-weighting or misbranding, maintenance of standards of cleanliness, and so on.

Thus the plan pays a basic minimum towards the dispensing charge, but the remainder of the charge is set by each pharmacy and charged to the patient.

For such a program to work, however, it is also necessary to ensure that the market-structure/conduct preconditions are met. If pharmaceutical associations promulgate uniform dispensing fees and can discipline members who price competitively, the direct patient charge is a useless device. Similarly, the accumulation of large numbers of pharmacies under the same or related management precludes price competition between chain members, although informal price collusion between independent pharmacies can have the same effect. More subtle forms of restraint would involve professionally enforced 'ethical' bans on advertising of dispensing charges, thus raising the search costs associated with consumer price selectivity. Nominal 'quality' regulations can serve the same purpose. Much of the potential gain in dispensing efficiency (and reduction in dispensing charges) would come through greater output per pharmacist, using many more dispensing assistants and spending less time in other retail or managerial functions. Professional restraints (in the name of quality) on auxiliary use or scale of operations can be used to prevent efficiency-enhancing/cost-cutting behaviour and so dilute or eliminate the gains from competition. Cady (1975) estimates that in the United States in 1970, legal restrictions and prohibitions on competitive practices in retail pharmacy, such as advertising of prescription drug prices, raised drug costs by \$220-242 million. Nor does his methodology include potential gains from a major expansion in the role of dispensing aides. Evans and Williamson (1978) estimate potential savings of \$35-50 million for a hypothetical universal pharmacare program in Ontario in 1975. One could move to a fully efficient (given present technology) dispensing system, but to reap these gains one must combine charges to patients with an energetic policy of promoting price competition between pharmacies; otherwise patient charges are mere ideological window-dressing, with no more practical effect than, say, a uniform \$1.00 charge per prescription.

As for the broader problems of prescribing appropriateness and costs of drug manufacturing, wholesaling, and marketing to physicians, these are important issues, but they are areas where patient choice is of little or no significance. Thus patient charges would be ineffectual there too. To impose a flat prescription charge as a device to limit general overprescription and overuse of drugs, for example, is to perform a symbolic act indicating that one recognizes the problem but chooses for political reasons not to address it. The appearance of action is often more politically expedient than either action or inaction.

In evaluating a plan such as the Saskatchewan one, we find the by now familiar results regarding the risk reduction and wealth transfer effects of a direct charge approach. Compared to a first-dollar coverage plan, a pharmaceutical insurance plan with differential dispensing charges set by individual dispensers clearly increases the financial risk borne by patients, while transferring wealth from users of drugs to taxpayers in general. The dispensing fee will result in either lower or unchanged levels of utilization, depending upon its deterrence power.⁹ In either event, to the extent that the dispensing fees create competitive pressure, lower-price pharmacies will be able to expand their market shares, thereby potentially lowering total prescription drug costs. Moreover, all pharmacies are placed under competitive pressure to improve the efficiency of the dispensing process.

Beyond the realm of prescription drugs, however, it is hard to find types of health services that would afford a useful role for differential charges. The work of Benham (1972) and Benham and Benham (1975) suggests strongly that, subject again to certain regulatory standards on lens quality, open advertising and price competition can significantly lower costs in the eyeglass and contact lens industries. But while not insignificant, these (like drugs) are peripheral to the 'leading sectors' of hospital and medical care.¹⁰ If significant steps towards attaining insurance and system objectives are to be made through direct charges to patients, both physicians' and hospital services must be included.

This suggests that we leave specific peripheral services and consider the implications of broader service repackaging. Under fee-for-service reimbursement, most proposals for patient participation in payments for health services are clearly designed to *increase* the flow of payments into health care and thus cannot be taken seriously as cost-containment methods. The root of the problem is that, with specific exceptions such as drug dispensing and eyeglasses, the present definition of the health care 'product' is not sufficiently standardized for the consumer to make informed, price-guided

9 Whether lowered utilization leads to a more effective pattern of utilization depends on the efficacy of the drug involved and on the effects the direct charges might have on patient compliance with drug regimens.

10 Provisional estimates from the Division of Health Economics and Statistics, Health and Welfare Canada, of 1975 health care costs in Canada are \$11.4 billion, of which prescription drugs and eyeglasses account for about 6.3 per cent. Hospital spending makes up 49.6 per cent and physician care 16.7 per cent, with physicians, of course, exercising a powerful influence over hospital and drug expense in addition to their direct costs.

consumption decisions. But it may be possible to repackage the product so that informed consumer choices are possible. To return to Arrow's (1963) original point, the complexity of a production process is irrelevant if the consumer can judge the value to himself of the final product. In health care, fee-for-service reimbursement requires the consumer to purchase the individual components of care for an illness episode, thus maximizing the information gap between the consumer and the technically informed provider. So long as the *individual* health care services are defined, priced, and sold as the 'final product', the consumer has no hope of making informed choices about the value of such services compared to the price he is requested to pay for them. In such circumstances price signals to consumers can play no constructive role.

Repackaging, however, would allow consumers to choose between more comprehensive care alternatives. Two concepts are essential in repackaging. First, the patient is expected to choose between delivery system alternatives, not between specific service or provider alternatives, and to respond to price information and incentives in this choice. Second, new delivery institutions must be created which combine insurance and service functions, such as groups of providers contracting with payors to provide comprehensive care at negotiated capitation rates and in effect competing with fee-for-service physicians.¹¹ In order to achieve cost containment, charges to patients must then be structured to channel patients to the lower-cost providers. The health maintenance organization (HMO) concept, for example, offers the patient a package of 'necessary health services' rather than specific items. At present, although Canadian consumers have 'free choice of provider', which implies that their decision to use more or less costly modes of service holds no financial implications for them, they are generally denied another freedom — that of choosing between this system and systems which would reward a choice of less costly modes of service. A constructive use of financial incentives might be to pay rebates or to lower premiums to groups of patients who choose service delivery organizations with lower costs. The Sault Ste Marie clinic, for example, was shown to use significantly less hospital care in serving its membership than was provided to a matched population receiving primary services from private practitioners (Hastings et al., 1973). Some of the resulting savings might be returned to users of this clinic. If the HMO idea

11 Whether one calls them community health centres, pre-paid group practices, closed panel plans, health service organizations (HSO), health maintenance organizations (HMO), or whatever is irrelevant.

generally leads to less hospital use, financial incentives to patients should be used to encourage its expansion.¹² Similarly, group dental insurance contracts could be written, not on an open-ended basis, but with groups of providers (closed panel) on a per-capita basis, thus providing incentives to the dentist groups to improve productivity by more use of auxiliaries and less provision of marginally beneficial or ineffectual services. Groups of providers could contract with government insurance plans, private insurance plans, unions, or companies, as long as the above concepts were embodied in the contracts. It should be obvious, however, that systems relying on competition between new kinds of provider organizations would require significant legislative changes to remove the present professional constraints on competitive behaviour and to prohibit collusive marketing practices in the name of self-regulation.

The recent revisions to the Canadian anti-combines legislation go only a very limited way in this direction, since they focus primarily on prevention of collusive price-setting. This has little relevance to health providers who bargain with government, and in any case may be circumvented by 'conscious parallelism', although a substantial increase in opting out by physicians might subject them to federal competition legislation. The new legislation does not, however, touch the critical area of self-regulation, the power to determine who shall be permitted to supply what types of services and through what organizational forms. That power controls competition from auxiliaries, enabling dentists, for example, to outlaw dental nurses by regulation except where, as in Saskatchewan, a separate Act establishes their occupation. Similarly, the type of flexible contractual arrangements envisaged above would not be possible within the existing regulatory framework. It is beyond the scope of this study to specify all the regulatory changes which might be necessary to ensure the development of delivery system alternatives. The intention here is to identify the potential for improvements in efficiency of health care delivery through such alternatives, and to indicate that certain types of direct charges to patients would be an integral part of a strategy for promoting and extending such alternatives.

Social acceptance of these delivery system alternatives would depend on their therapeutic equivalence being established on a continuing basis by external monitoring or public regulation, because doing so is beyond the capability of patients in any system yet devised. Various institutional arrange-

12 Although much has recently been written on this theme, the central issues and evidence are summarized in Barer (1977, 1979).

ments could achieve this. In dental care, for example, where measurement of the quality of results is relatively straightforward, one could conceive of group capitation contracts for dental services that specify the quality levels to be maintained. A separate contractor could fairly easily determine whether these standards had been met, through patient sampling and the application of agreed-upon criteria. In the more complex field of medical and hospital care, where outcome measures are more ambiguous, some form of external public quality monitoring agency might be necessary, with inspection teams like bank inspectors.

It has been suggested in the United States that service repackaging (combining service and insurance functions in a single organization) and public quality monitoring may not be necessary to resurrect competitive efficiency. Havighurst (1977), for example, has argued that private insurers might combine insurance and monitoring functions and exert direct pressure on providers to improve efficiency in order to maintain the competitive status of the insurer's premium — in effect private insurance companies might act as informed consumers. The historical precedent for this, described by Goldberg and Greenberg (1977), seems, however, to show that while such monitoring is technically feasible, it is impossible in practice in the face of co-ordinated opposition by providers.¹³ That any legal framework could be designed or achieved to eliminate co-ordinated opposition is doubtful.

In the Canadian case, at least for medicine, the private insurer concept is less relevant. Furthermore, both unions and companies are likely to experience similar difficulties in dealing with the co-ordinated opposition of providers. Consequently, any significant repackaging in the Canadian con-

13 An apparently successful program of provider-monitoring by private insurers in Oregon during the early part of this century was terminated in the late 1930s and early 1940s when a physician-sponsored insurance plan drove the companies from the market. Once their own plan was established, physicians refused to deal with any private insurer attempting to monitor their behaviour. In 1971, the Aetna Insurance Company, one of the largest health insurers in America, introduced a relatively limited policy to try to constrain over servicing by physicians. The policy was rapidly terminated in the face of energetic and co-ordinated physician resistance. More recently, Michigan Blue Cross is attempting, through the reimbursement process, to discourage excessive use of diagnostic tests. The policy was introduced on 1 January 1978 over strong medical society protests, and no results are yet available. What is significant in the Michigan case is that the insurer's action is backed up by a small group of large insurance purchasers (the UAW and the auto companies), which may strengthen its position ('Michigan's clamp on doctors bills', *Business Week*, 19 December 1977, 34).

text appears to require assumption of the monitoring/regulation role by the government.

Even with such new institutions, direct charges will, as usual, increase financial risk for patients and transfer wealth from users of care to taxpayers in general; however, to the extent that such institutions provide incentives for more effective and efficient patterns of care, cost escalation would be mitigated.¹⁴ Whether or not consumers will persist, for reasons not related to health, in choosing higher out-of-pocket cost system alternatives over lower-cost alternatives of equal accessibility and therapeutic effectiveness, and if so whether or not this is cause for public concern, is a separate issue, to be dealt with in the following chapter.

From the discussion in this chapter it is clear that routes are available for the implementation of provider-determined differential charges; however, it is questionable that such direct charges to patients could satisfy the market structure/conduct and product-characteristic/consumer-information preconditions without the intervention of some third party.¹⁵ For this reason we now consider a final category of schemes which, while maintaining the differential nature of the direct charge across providers, rely upon the participation of a third party, normally a government agency, to create the differential and monitor quality across the delivery alternatives.

14 This of course assumes no significant feedback effects of fee-for-service providers in response to reduced market shares. Since physicians not practising out of HMOs must lose patients to those institutions if the financial incentives are successful, they may respond by increasing the intensity of care provided to their remaining patients. If HMOs began to attract a significant segment of the population, the impact of that feedback would be mitigated, so that implementation of this type of 'direct charge' on a wide scale would be subject to considerably less potential overriding feedback than any of the uniform charges considered.

15 From the discussion thus far it should be clear that modifications to market structure, i.e. limiting or removing the power of professional associations to regulate the economic conduct of their members, will be necessary but not sufficient to induce price-competitive behaviour. Positive intervention to influence the conduct of providers by promoting competitive behaviour and prohibiting collusion will also be necessary. If this behaviour cannot be induced, or is considered socially undesirable, then direct charges serve no useful function.

5

Non-provider determined differential charges

A market is like a tool: designed to do certain jobs but unsuited for others. Not wholly familiar with what it can do, people often leave it lying in the drawer when they could use it. But then, they also use it when they should not, like an amateur craftsman who carelessly uses his chisel as a screwdriver.

C.E. Lindblom (1977)

Although some of the preceding schemes involving differential charges across providers might improve health system efficiency, they are beset by two fundamental problems. In the first place both formal and informal mechanisms exist to prevent effective price competition; furthermore, even if such competition were to occur, there is nothing inherent in a system of provider-determined charges to ensure that differentials are a function of quality and/or efficiency differences. Since the single outstanding characteristic of the health care 'market' is the inability of the consumer to judge accurately either the need for services or the quality of the same services, it is clear that the creation of any efficient market process will necessitate intervention by some third party to provide an adequate level of accurate information.

If this were the only condition to be satisfied, the non-provider intervention might be accomplished by insurance companies, firms, or unions of an appropriate size and inclination.¹ The existence of other factors, however,

1 Again, such optimism is subject to the lessons from Aetna (Goldberg and Greenberg, 1977).

suggests that it is reasonable to restrict the current discussion to intervention by governments, although in principle the product-characteristic/consumer-information precondition could be satisfied by private intervention even under schemes requiring public intervention on other grounds. But government intervention will almost certainly be necessary to satisfy the market structure/ conduct precondition, because only if there are no (successful) restrictive practices will differential charges influence health system efficiency. Furthermore, government intervention would appear to be the appropriate form of third-party intervention in an industry already dominated by public health insurance. Finally, as noted in the preceding chapter, it is not clear that we would wish, as a society, to ration access to higher-quality care by ability to pay even if price differentials did accurately reflect quality. Thus, government intervention may be required to achieve the social objective of 'equal access' embodied in the health insurance legislation.

The phrase 'non-provider determined' needs further clarification. It is not meant to imply that the government would determine the level of every charge for every provider in a differential charge system; such a task is surely neither operationally feasible nor socially desirable. Rather, it means that as a result of government intervention and policy a differential constituting a direct financial incentive or disincentive to patients is created between two or more types of providers or delivery settings.² As will soon become apparent, there may still be substantial scope for provider-determined charges within one of the provider or delivery options.

This chapter continues and completes the categorization of proposals for direct charges, illustrated again in Figure 8. But it differs from earlier chapters in one important respect. Whereas they examined well-defined or well-publicized schemes, the proposals considered here might best be thought of as potential policy directions warranting energetic further exploration. Both avenues suggested below, 'selective deinsurance' and 'parallel systems', await further clinical information and conclusive economic evaluation. Consequently the following analysis is intended only to develop the rationale for

2 This illustrates the fact that no taxonomy can ever be watertight. Under this broad definition of non-provider determination, some examples of service repackaging outlined in the previous chapter might just as easily be considered here. Similarly, as noted below, the initial policy direction considered in this chapter might have been dealt with in an earlier section on per-service charges. As has been emphasized on several occasions, however, the uniform vs differential and provider vs non-provider determination distinctions seem to be of sufficient analytic importance to serve as taxonomic cornerstones in what has all too often been a confusing debate.

FIGURE 8

Methods of increasing patient participation in payment for health care:
non-provider-determined differential charges

		Charges determined by	
		Non-provider	Provider
Level of charges across providers	Uniform	Coinsurance Deductibles Per-service charges Income- and income tax- linked proposals	Coinsurance Per-service charges
	Differential	Selective deinsurance Parallel systems	Major risk medical Extra-billing Service repackaging

each policy direction, to identify obstacles to operationalizing the proposals, and to make some preliminary estimate of their potential economic impact.

SELECTIVE DEINSURANCE

A policy of selective deinsurance is defined to be third-party (in this case government) refusal to reimburse patients or providers for inefficacious services received or rendered. Such a policy might be applied to the subset of currently offered and insured services that have been demonstrated to be of little or no health benefit. On the premise that it is inappropriate for public health insurance to cover procedures shown to have no clinical effectiveness, such procedures would be 'deinsured', thereby permitting providers to bill patients directly for utilization of them. The direct charge to patients would then become 100 per cent of the provider's fee. While numerous difficulties are immediately apparent, deinsurance precedents do exist. Certain cosmetic surgery procedures and physical examinations for insurance or other certification purposes are normally omitted from public insurance coverage, pre-

sumably on the ground that they do not contribute to or promote health.³ Similarly, certain drug plans refuse to reimburse the cost of drugs or combinations of drugs shown to be inefficacious or to be no more efficacious than significantly less costly alternatives.

While the principle underlying selective deinsurance may well meet with general agreement, both the operational feasibility and the economic impact of the approach are questionable. The continuing information requirements would be extremely heavy, and their cost may be prohibitive in light of the few procedures, if any, which at present appear to meet the criterion for deinsurance. Such a criterion would involve rigorous demonstration, presumably in multi-centre, controlled, randomized trials, that a given service or procedure exhibited either no health benefit or positive harm compared to the next best alternative (which might be to do nothing). In addition to the need for a multitude of trials to take account of all variants of procedures and new interventions, this approach is significantly undermined by one common observation: the main problems of unnecessary servicing are not in the area of demonstrably useless services. There is rather a large grey area of procedures that may do some good to some patients. The fact that tonsillectomies are only indicated in between 5 and 20 per cent of the cases for which they are performed⁴ does not imply that the procedure is wholly without merit, and one would have trouble justifying 'deinsurance' in all cases. Partial coverage, with the coinsurance rate varying with the proportion of cases indicated, would not be a solution, overcharging all those who needed the procedure and undercharging the rest. It would fail to discourage the unnecessary procedures, because neither the reimbursor nor the patient would know which these are. The physician is the only one potentially capable of making this distinction, and a partial coverage approach does not put him at risk.

Topical fluoride application in communities with fluoridated water presents the same problems. As a clinical procedure its benefits appear to be far too small to justify the costs of application, and one might therefore wish to

- 3 It is not clear that one could ever deem any procedure to be inefficacious if the definition of health is made broad enough to encompass mental health and anxiety dimensions. Cosmetic surgery might, indeed, be difficult to deinsure if health includes mental health and its derivatives, among them mirror-gazing.
- 4 See Roos et al. (1977a). The authors apply alternative standards to the pattern of illness episodes which should be observed before a tonsillectomy is justified. The loosest criteria, or the least stringent preconditions, for tonsillectomy were met in only 18.6 per cent of operations performed.

refuse reimbursement. But there may still be individuals with 'rampant caries' in such environments for whom the procedure is indicated. How does one target these individuals for benefits without supporting universal application of a procedure with dubious merit?⁵

- 5 The determination of the cost-effectiveness of topical fluoride treatments has been hampered by the difficulty of establishing efficacy levels. Davies (1973) provides an early attempt, while Forester and Schultz (1974) provide a survey of the state of the art at that date. A number of studies are also surveyed in the *Report of the Children's Dental Health Research Project* (British Columbia, 1975, chap. 10). Since that time new results have been reported, but the general efficacy issues have yet to be settled.

It appears generally accepted that topical fluorides in experimental settings can reduce the incidence of dental caries by anywhere from 25 to 40 per cent. Higher and lower estimates exist, and there are many different agents and techniques of application. An average rate of caries incidence (as opposed to prevalence, which can be much higher in untreated populations) is between one and two teeth a year in childhood, so a very rough estimate would be that topical fluoride application for, say, one hundred children might avert decay of a maximum of fifty to eighty teeth over the next year. If the children live in a community with fluoridated drinking water, however, the percentage reduction in caries incidence as a result of the topical applications is smaller, and some studies find no effect (Forester and Schultz). What must be recalled in interpreting such results is that even if the percentage decrease from topical fluoride application were insignificant, the base incidence rate falls sharply. Water fluoridation cuts the incidence rate by half or more. Thus if the percentage reduction in fluoridated communities were a very optimistic 20 to 30 per cent, the base incidence expected would be only one half to one carious tooth per child per year. Treatment of one hundred children would thus save at most only twenty to thirty teeth, compared with fifty to eighty in the unfluoridated case. Reductions of 20 to 30 per cent would be very optimistic for a fluoridated area, while a caries incidence rate of one tooth a child in the absence of topical fluoride is quite pessimistic.

Moreover, the reported studies are themselves upper-bound estimates. They generally study children in the 9-11 age range where the rate of uptake of fluoride ion is greatest. Compliance problems are eliminated, and the application technique is controlled under experimental conditions. No one has any information as to the extent of deterioration of efficacy under field conditions in individual dental practices, but some researchers suspect it may be substantial.

Finally even if efficacy could be assured, the cost-benefit calculations are not straightforward. One cannot simply compare the cost of one hundred topical fluoride treatments (and associated tooth cleanings and polishings) with that of restoring fifty, eighty, twenty, or thirty teeth (assuming, say, an average of 1.75 or 2.0 decayed surfaces per tooth). On the one hand prevention has non-monetary benefits because restoration hurts and because a sound tooth may be aesthetically superior to a restored tooth. On the other hand a single application may merely defer decay for a year. A properly restored tooth may not redecay for ten years, while the

Instead of unequivocally deinsuring procedures or services, a more sensitive approach might be to identify procedures whose benefits do not justify their performance under certain well-defined circumstances or beyond certain frequencies, followed by deinsurance accordingly. Pap smears and well-baby visits are examples of procedures whose marginal effectiveness in current patterns of utilization is suspected to be zero for some individuals.⁶ A wide range of additional suspect procedures can be gleaned from the medical literature. Again, however, the informational requirements for confirming the inefficacy of suspect procedures are immense, and the range of identifiable deinsurance candidates is at present limited. Moreover, granting for the moment that the necessary trials could be done and did turn up enough candidates to have some economic results, at least two obstacles remain.

First, the conditions of deinsurance would have to be communicated to consumers in such a way as to indicate clearly the circumstances and frequencies in or beyond which they would henceforth be at financial risk. Such communication would also have to convince consumers of the accuracy of the evidence. Failure to do so would result not in a decrease in utilization of deinsured services but simply in an alteration of the relative private and public shares of the total cost of care. Second, provider co-operation would be crucial to the success of such a policy. In fact, if the government's communication of clinical evidence to prospective patients was totally successful, one might argue that there would be no need to deinsure procedures other than to prevent continued prescription of them by providers. Whether for economic reasons or because of differences of opinion regarding the therapeutic value of deinsured procedures, provider reaction would still represent

'prevented' caries may emerge in elevated incidence next year unless continued reapplications occur. The preventive strategy then becomes much more costly compared to its benefits.

The type of longitudinal efficacy data necessary to evaluate the cost-efficiency of dental prevention does not exist, even experimentally; the field data do not exist at all. But it seems safe to say that if there is any benefit from topical fluoride applications to children or adults in fluoridated communities, the orders of magnitude are probably around one tooth restoration avoided for each ten topical fluoride treatments and cleanings (give or take a tooth), and even that benefit may simply be a deferred restoration, not an avoided one. The cost of each treatment varies by fee schedule and region, but generally the fee for a prophylaxis and fluoride treatment would be about the same as for the restoration of a two-surface decayed tooth. The ratio of costs to benefits is about 10:1.

- 6 On the former see, for example, the report of the federal task force on screening for cervical cancer (Walton et al., 1976). Hoeckelman (1975) and Yankauer (1973) comment on the latter.

an uncertain variable in the analysis. Potential reactions, all of which would adversely affect any deinsurance policy, include public and/or provider questioning of the government's clinical evidence, provider reclassification of procedures where possible (e.g. from periodic exam to some symptom-related procedure) to avoid those deinsured, and temporal shifting of the treatment episode to enable performance of the desired procedures at no cost to patients. In addition, aggregate supplier response to a policy which essentially lowers provider incomes as well as questioning existing practice patterns is not likely to assist cost control efforts.

With respect to our four public policy goals, definitive assessments are precluded by the tentative nature of the policy direction and the inadequate information on the range of procedures which might be affected. It does seem reasonable to suggest, however, that a policy of selective deinsurance would be approximately neutral with respect to both risk reduction and wealth transfer. By definition, any increases in risk or actual expenditures would stem from the actions of informed consumers making choices on non-health-related consumption goods, a situation which not only has little to do with health insurance or system objectives but also is one in which we are normally content to allow markets their usual, if imperfect, sway.⁷ There is little one can say about differences in use by income class of 'deinsurable' procedures without first having a well-defined and supported vector of such procedures. However, we see no obvious reason for the incidence of such procedures to be a significant function of income class. The most one can say is that there will be a wealth transfer from those who continue to use deinsured services to taxpayers in general. Although the effects of deinsurance

7 One might argue, to the contrary, that 'deinsurance' of particular services implies that users of those services must pay for them out of pocket. Insofar as use of such services is, like any health service, unpredictable, the consumer/patient is exposed to increased risk by this policy. Such an argument appears to us to neglect the reasons for a service being 'deinsured'. Either some therapeutically equivalent and less costly alternative insured service is available, or the service in question has been shown to be inefficacious. Accordingly the decision to utilize the 'deinsured' service is a choice made on grounds unrelated to health and cannot be thought of as necessitated by some unpredictable illness or accident. Failure to use the service would have no impact on health status, either because the service was inefficacious or because the insured alternative was used instead. Thus, the patient/consumer is not 'at risk' and can freely accept or avoid the expenditure without threat to health. Of course such a position requires that the patient possess the appropriate *information* to make the utilization decision, but 'deinsurance' is also a relatively powerful way of conveying that information.

on efficiency and aggregate utilization depend largely on assumptions regarding provider behaviour, the incentives to technical efficiency through effective patterns of utilization are strong and direct.

PARALLEL SYSTEMS

A second and perhaps more promising avenue through which government may attempt to influence efficiency directly involves programs which create parallel service systems in which direct charges to patients reflect differences in technical efficiency. Proposals in this category stem from the observation that for certain services the same therapeutic outcome may be attainable by at least two production processes involving different input mixes. To the extent that the different input mixes generate differing production costs for the same outcome, scope exists for price signals, in the form of differential direct charges to patients, to encourage competitive behaviour and thereby technical efficiency.⁸ For example, if identical immunization services can be provided through either family physicians in their offices or nurses in public health clinics at a differential cost in favour of the clinics, a case may be made for providing the services 'free' through the clinics. Family physicians would be permitted to bill patients directly for immunization services provided in their offices, thereby making patients fully responsible for such bills. This is equivalent to deinsuring the physician-supplied service if a lower-cost, equal-outcome alternative is implemented. Although this example involves a public/private choice, the ward/semiprivate/private price differential for hospital space under existing public hospital insurance is another example of a differential direct charge allowing informed consumer choice.

In effect, both these cases amount to basing differential charges on ambience, a dimension in which informed consumer choices can reasonably guide resource allocation. In fact, extension of this approach could have significant potential for restraining health costs — a universal system of public health unit well-baby care, for example, or a school-based children's dental care service, could coexist with private practice systems in which physicians

8 The subsequent examples and discussion assume that the lower-cost technology will be exploited by a public program. One might ask why private suppliers of health care services, whether self-employed or not-for-profit, would not themselves adopt technologies providing a given therapeutic outcome at minimum cost. For a survey of evidence and an analytic explanation showing that self-employed proprietors will generally not employ least-cost technology, see Evans and Williamson (1978, chap. 6).

or dentists charged patients directly for the service. The key to the use of such parallel systems, however, is that there must be a therapeutically equivalent and accessible public service available for which patients are not charged. If this alternative is not available, and providers are allowed to form their own judgments of the worth of their own services, we are back to the original case of independent provider determination of charges. Since patients are generally unable to distinguish between quality and ambience,⁹ and since providers obviously have no incentive to assist them in this distinction (e.g. by pointing out that some other provider is as competent and cheaper but a less gifted conversationalist!) the theoretical consumer sovereignty argument has no practical force.

Another setting in which we might envisage consumers being able to make informed choices about the purchase of a clearly defined product, but which may not require a public/private set of alternatives, is suggested by the work of Evans and Robinson (1973, plus preliminary results from a followup study) on day-care surgery. In this case patients have the choice of resolving a well-defined episode of illness requiring surgery as conventional in-patients or as day-care patients. Since per-episode cost savings of up to 70 per cent are occurring in the out-patient mode this seems a technically superior alternative. Offering an incentive to the patient which will encourage choice of the lower-cost, therapeutically equivalent option seems a not unreasonable deployment of the direct charge concept. For example, patients who choose the in-patient route could be charged some proportion of the differential costs. Of course any efficiency gains and gains with respect to our utilization objective are again at least partially dependent on a presumed absence of secondary feedback effects. The beds freed by the day-care surgery unit would have to be eliminated if cost savings were to materialize. Furthermore, sufficient day-care facilities would have to be made available to handle all cases potentially treatable in that setting, and patients would have to be informed of their existence, of the incentive involved, and of the equal therapeutic impacts. If patients are influenced in their choice by, for example, hospitals with incentives to encourage the in-patient route, the fully-informed-consumer-choice assumption will not be satisfied. While this type of parallel system remains attractive even in that context, we cannot rely on the price nexus to channel patients to the lower-cost alternative.

9 What supports the argument for direct charges in pharmaceutical dispensing is precisely that the quality and ambience dimensions can be disentangled; the consumer can make (and, equally important, perceive himself to make) decisions in response to relative prices in the latter dimension which are independent of the former.

The need for demonstration and communication of the therapeutic equivalence of any parallel system cannot be overemphasized. Any doubt allowed to exist regarding potential quality differences between the systems will only lend substance to the inevitable charges of 'two-class medicine'. Furthermore, an active public information campaign will be required, since established providers will have every incentive to undermine the public alternative both through the statements they make to their patients regarding its quality and through political activity directed at restricting the scope and budget of the public program. If these political attacks cannot be met, the parallel system option will fail.

A preliminary assessment of the impact of these parallel system examples in light of our policy objectives suggests results very similar to those of selective deinsurance. The proposals would be neutral with respect to risk reduction and wealth transfer for those using the public alternative, while for those choosing the private alternative financial risk would be increased and transfers would occur from users to taxpayers in general. With respect to levels of utilization, and considering the nature of immunization services, well-baby care, and children's dental care, it does not seem unreasonable to suggest that a vigorous advertising campaign alone may increase demand, especially when reinforced in the case of dental care by a new financial incentive to parents. With regard to efficiency and effective patterns of utilization, while the public alternative provides the scope and conditions for improvements on both counts, the possibility of countervailing shifts in behaviour by private providers clouds any attempt to estimate the effects of parallel systems on aggregate (i.e. public plus private) expenditures or efficiency.

The parallel system approach contains one further variant warranting scrutiny, which might be labelled 'incentives for self-care'. Instead of employing direct charges to deter inappropriate patterns and levels of utilization, this approach consists of providing positive financial incentives, e.g. stipends, to patients undertaking their own care (with supervision) at home rather than in a hospital setting. Analytically, the approach is similar to that of parallel systems in that the differential costs of producing care stem from differences in input mix. If, for a specific service, a 'patient-intensive' process demonstrates cost savings relative to a 'professional/ institutional-intensive' process which produces the same health outcome, there may be scope for returning some proportion of the savings to patients to encourage increased utilization of the less costly process. Operationally, it is similar to service repackaging, in which cost savings are returned to users of health service organizations if such organizations demonstrate lower costs than established fee-for-service

alternatives. The particular mechanism employed to return the savings is not considered here, although a variety of potential mechanisms such as cash rebates, lowered premiums, stipends, and tax credits exist.

Although a systematic analysis of opportunities for implementing this approach has not been made, an obvious area in which to begin such a policy is chronic care. Consider, for example, the extremely expensive and increasingly important area of kidney dialysis. While hospital dialysis is mandatory for some dialysis patients with certain co-morbid conditions or a high risk of complications, large numbers of dialysis patients without contraindications for home dialysis continue to receive hospital treatment.¹⁰ From an individual patient or family viewpoint, this is an entirely rational decision, since home dialysis at the very least involves a significant amount of effort on the part of the patient and his family, and can sometimes entail significant disruptions in family functioning. On the other hand the public savings associated with home dialysis are significant. Recent studies indicate that, even allowing for the training and startup costs of equipping households with dialysis machines, savings may be on the order of \$10,000 to \$15,000 a patient a year (Friedman et al., 1978; Bennett and Krasny, 1977). Under these circumstances, there would appear to be ample scope for earmarking some portion of this savings as an incentive to encourage patients to accept home dialysis.

This approach need not be restricted to the chronic care field, although that would appear to be the one area with the highest marginal gain. For example, evidence suggests substantial cost savings in a program of home administration of intravenous antibiotics by patients.¹¹ In both of these cases, however, and in any situations under consideration for the 'incentives to self-care' approach, the demonstration of cost differentials is of little

10 In fact, Rennie (1978) reports a significant *reduction* in the proportion of all chronic dialysis treatments performed in the home in the United States between 1972 and 1976, apparently reflecting the ambience advantages to the patients and/or economic advantages to the practitioner of hospital dialysis and the inadequate encouragement of home dialysis. In Canada, on the other hand, the percentage of dialysis patients receiving treatments outside the hospital has increased from 23.3 to 26.9 per cent over the same period (Canadian Renal Failure Registry, personal communication with Dr Arthur Shimizu).

11 A Manitoba study, in which patients with problems such as staphylococcal osteomyelitis, bacteremia, and infective endocarditis carried out their intravenous antibiotic therapy at home, reported average cost of treatment per patient day as \$39, compared to \$137 per day for similar treatment in hospital (*Canadian Medical Association Journal*, 118, 18 Feb. 1978, 445).

policy significance unless the clinical effectiveness of the home alternative can be shown to be no worse than that of the institutional one. While these conditions would require more conclusive evidence than is currently available in most instances, the means for obtaining such evidence (clinical trials, cost-effectiveness analyses) are well established and are being increasingly employed. As usual, the more difficult tasks will be to communicate effectively the foundation for and conditions of the program to prospective participants and secure the co-operation of the providers involved. To the extent that incentives for self-care focus on procedure-specific interventions and well-defined patient populations, these tasks may prove somewhat easier in this case than in the selective deinsurance or parallel system examples discussed above. With regard to the economic impact of this approach to direct 'charges', it should by now be apparent that no one route is likely to provide instant relief from cost escalation; however, if applied selectively, incentives to self-care have a potential for significant cost savings because they address the most costly types of interventions, those involving hospitalization. In addition, they are attractive on grounds of utilization and efficiency. There will be an obvious incentive to shift the pattern of service utilization from the acute care hospital to the private home. Given equal outcomes from the two alternative treatment locales, adoption of the home-care option will certainly be a step toward more efficient production of health services. 'Incentives for self-care' will be neutral with respect to risk effects, and it is not clear what, if any, impact there will be on wealth transfers. There is clearly a transfer of wealth, for example, from taxpayers to uremia patients who receive the incentive payment, but the distribution of such patients by income class is unknown.

This chapter has been admittedly exploratory in nature. The selected examples chosen to illustrate the potential scope for deinsurance, parallel systems, and incentives to self-care were just that — examples. We have not tried to be exhaustive because the conceptualization and definition of informational requirements seem at this stage far more important than any attempt to establish a comprehensive list of potential procedure and illness candidates.

6

On avoiding future snares: summary and policy conclusions

If years of ignorant assumption have convinced [American] policy-makers that cost-sharing can reduce unnecessary utilization and/or hold down social insurance costs, it will require more than impassioned rhetoric on behalf of the disenfranchised poor to reverse the pattern.

C.P. Hall (1974)

The stated purpose of this paper was to explore, under quite general assumptions, whether direct charges to patients offered significant tangible social benefits. Unfortunately for public policy (because such charges are often relatively easily implemented) the answer seems to be no. In fact, most such policies appear likely to produce positive social harm. It has been shown that the necessary preconditions for the success of direct charges in achieving a plausible range of policy objectives are rarely satisfied and, that when they are, other operational impediments may preclude policy success.

The taxonomy chosen to categorize the various possible direct charge schemes emphasized the distinction between charges which vary across providers and those which will be identical regardless of a patient's choice of provider. A second dimension of differentiation identified the party responsible for determining the level of those charges. This scheme, illustrated in Figure 1 and carried with progressive additions through the report, allowed the placement of each direct charge scheme considered into one

of four categories – uniform charges determined by third parties (governments or insuring agencies), uniform charges established by individual practitioners (or their practices) or professional associations, and a similar division of differential charges. The policy objectives presented in chapter 2 were chosen as appropriate pursuits of public programs entrusted with providing health care insurance and/or services. They were divided into insurance and system objectives. The insurance objectives were concerned with the impact of each type of charge on consumer risk-bearing and on the transfer of wealth between and within income classes. In contrast, the health system objectives abstracted from risk and wealth transfer effects to focus on the mix and level of services provided by the health care delivery system and the technical efficiency with which they were provided. The latter objective was defined in terms of producing a given mix and level of services at minimum real resource cost. The utilization mix and level objectives took this one step further by suggesting that, while we ought to be interested in producing a given service at minimum cost, we should not be interested in producing inefficacious services at any cost. Closely related to these system objectives was the impact of different direct charges on health care costs. Of course once the qualitative (and in some cases quantitative) impact of a charge on utilization and efficiency has been determined, the likely impact on total health expenditure follows directly.

Drawing on that taxonomy, we proceeded in chapters 3 through 5 to fill out Figure 1 with specific schemes, evaluating each in turn against these objectives under a set of plausible assumptions about the reactions of the various parties involved. Table 2 summarizes the results in those chapters by indicating the expected impact of each direct charge proposal in terms of each of the four objectives outlined in chapter 2. The contents of that table may strike the careful reader as being considerably more definitive than is warranted by the discussion in those earlier chapters. It is worth emphasizing, then, that the symbols of Table 2 are ‘best guesses’, subject to the qualifications specified not only in the notes accompanying the table but also in the more detailed discussions of those chapters.

Chapter 3 focused on uniform charges, both provider and non-provider determined. Coinsurance, deductibles, per-service charges, and income- and income tax-linked schemes were considered, with per-service charges and coinsurance receiving attention in both provider and non-provider determined contexts. Chapter 4 dealt with one half of the differential charge dichotomy, provider-determined charges, and chapter 5 was devoted to a discussion of some more innovative, as yet not fully defined schemes involving non-provider determination of differential charges.

As Table 2 indicates, any direct charge involving out-of-pocket, unpredictable expenses for the consumer will by definition increase the share of risk borne by the consumer compared to that in a first-dollar universal scheme (in which of course the risk is zero). Only in the bottom four schemes, where the patient is offered a no-risk, therapeutically equivalent alternative, or where the risks involved have nothing to do with health status (and thus are not related to unpredictable health-status-altering events), do we see non-negative impacts on the risk reduction objective. Similarly, charges whose aggregate levels for a given family are direct functions of utilization only will involve perverse wealth transfers – from the ill to the healthy and, to the extent that the poor (including a significant share of the aged) are less healthy than the rich, from low- to high-income classes. Once we link charges to both utilization and income, the relative wealth transfer effects of each linkage are difficult to determine (see note 2 of Table 2).

While one could assume that physicians will extra-bill according to their perception of each patient's ability to pay, thus eliminating part of the low-income to high-income wealth transfer for that segment of the population that is ill, extra-billing still involves a transfer of wealth from the ill to the well. The uncertain wealth transfer effect indicated for extra-billing in Table 2 derives from our closer look at the potential justifications for such extra-billing. We argued that just as direct charges for services which are not health-related entail no increase in consumer risk-bearing, so those charges also will not affect the distribution of wealth in a direction about which health policy planners ought to be concerned. Thus, higher-income patients may be more willing to pay extra charges for ambience effects, and since the 'services' rendered under that label are not related to health outcomes we cannot consider this a socially undesirable transfer of wealth from the sick to the well. It simply has nothing at all to do with illness. On balance, then, our question mark in Table 2 for wealth transfer effects of extra-billing may be generous. In all likelihood, the net effect of the various extra-billing 'justifications' on this objective will be negative. The question mark is motivated only by the fact that the wealth transfer effect depends on which of these 'justifications' we are considering.

Service repackaging also has uncertain wealth transfer effects but for a different reason. Under this heading we considered a variety of schemes ranging from charging patients for the dispensing component of pharmaceutical prices to wider-reaching repackaging involving the creation of products (groups of services) about and among which consumers can make relatively informed choices. While dispensing fee charges will still transfer wealth from the ill to the well, the net wealth transfer effects of rebates

TABLE 2

Impact of direct charge schemes in terms of policy objectives

	Risk reduction	Wealth transfer	Utilization Levels	Efficiency Patterns	
Uniform charges					
Coinurance	—	—	N	—	N
Deductibles	—	—	N	—	N
Per-service charges ¹	—	—	N	—	N
Income-linked charges ²	—	?	N	?	N
Income tax-linked schemes ²	—	?	N	?	N
Income-linked rebates ²	—	?	N	?	N
Differential charges					
Major-risk medical ³	—	—	N	—	N
Extra-billing ⁴	—	?	N	?	N
Service repackaging ⁵	N	?	+	+	+
Selective deinsurance	N	N	+	+	+
Parallel systems	N	N	+	+	+
Incentives for self-care	N	+	+	+	+

Note: The base of comparison is the currently operative universal first-dollar insurance program. + indicates a socially desired effect. For the risk objective this would entail a reduction in consumer risk-bearing; for wealth transfer a progressive shift; for patterns and levels of utilization a decline in inefficacious or marginally useful procedures; for efficiency the attainment of a given therapeutic outcome at lower real cost or improved outcomes at the same cost. — indicates socially perverse effect. N indicates no significant effect. ? indicates unknown effect.

- 1 For chronic or long-term care facilities where the occupant is a pensioner for whom the facility is domiciliary as well as treatment-oriented, a per-diem charge based on pension receipts may change risk reduction to 'N' and wealth transfer to '+'.
- 2 Wealth transfer effects across income classes depend on the strength of the relation between charges and income and on the progressivity of tax sources financing insured care. Effects within classes are '—'. Impact on patterns of utilization, if any, will depend on the variation across income classes in deterrent parameters chosen.
- 3 If the limit on patient liability is income-linked, the negative wealth transfer effect may be mitigated as in note 2.
- 4 The wealth transfer effects considered in this study have been those across and within income classes of patients. There is, of course, also a wealth transfer from patients to health care providers to the extent that extra-billing merely raises prices without affecting utilization. In addition, if providers respond to changes in income by altering technical efficiency, then the neutrality indicated in the final column will be changed.
- 5 Repackaging in the form of capitation contracts for the purchase of services and insurance enables the consumer to avoid risk, although different consumer groups purchasing different service/insurance packages may pay different premiums. In the

or lowered premiums associated with choice of low-cost modes of service (such as prepaid groups) will depend on the income class mix of those making that choice. If the population receiving care from a low-cost HMO is composed primarily of high-income families, it is clear that the perverse wealth transfer effects remain.

It was suggested that selective deinsurance would be approximately neutral with respect to wealth transfers because there was no reason to believe that potentially deinsurable services are received with greater frequency by low-income families. In any event, since all costs incurred after deinsurance would result from informed decision-making where a costless therapeutically equivalent alternative existed, any resulting wealth transfers are beyond the concern of this study. A similar argument holds for parallel systems.

It is only with self-care incentives that we might find positive wealth transfers resulting from direct charges, although as noted in the previous chapter this will be dependent on the income class mix of those receiving the incentive payments. Since the poor are generally of lower health status than upper-income groups, it is reasonable to assume that the opportunities for choosing the self-care option would be more common within that group. There would in that case be a transfer of wealth from taxpayers as a whole to the lower-income groups.

Turning to effects on levels of utilization, Table 2 suggests a division of schemes identical to that for the risk reduction objective. In this case we have argued that since provider incomes are a function of utilization levels, feedback effects will be set in motion which approximately negate any initial-deterrence-related reduction in those levels resulting from direct charges. Considerable empirical evidence was also offered in support of that argument. Once one moves into the realm of schemes allowing informed consumer choice between different-cost, equal-efficacy alternatives, however, it may become more difficult for those providers not involved in the low-cost options to react to reduced levels of utilization. Levels and patterns of utilization both will be favourably affected by these schemes. While physicians will undoubtedly be able to make adjustments elsewhere in response to the deinsuring (and subsequent reduced incidence) of any one service, the options for the high-cost pharmacist, for example, are less clear. Similarly, if an increasing share of the population opts for the HMO model,

case of pharmaceutical dispensing fees, however, individual patients are placed at some risk compared to a first-dollar insurance plan. The positive benefits from service repackaging require that the product-characteristic/consumer-information and market-structure/conduct preconditions be satisfied.

those practitioners on the 'outside' may find it increasingly difficult to maintain adequate income levels.

The impact of coinsurance, deductibles, per-service charges, and MRM schemes on patterns of utilization is clearly perverse. There is no evidence to suggest that consumption of the least necessary services will be deterred. The impact of income-linked schemes is difficult to determine (see note 2, Table 2), and the effect of extra-billing is uncertain because, again, numerous possible justifications were offered for it.

Uniform-charge schemes as a group preclude any impact on technical efficiency because by definition they involve charges which are identical for all providers. Since consumers then have no incentive to patronize one provider over another for cost-related reasons, providers will not be under any competitive pressure to increase their technical efficiency in order to reduce costs and prices. Consequently, whatever inefficiency may characterize the present delivery system will be maintained in the face of uniform direct charges. Even under such differential charge schemes as major-risk medical insurance and extra-billing, however, it was shown that there is no reason to believe that price differentials by themselves will foster the competitive pressure necessary to induce efficiency improvements. Other consumer information and market structure conditions must also be satisfied. Moreover, to the extent that major-risk medical and extra-billing schemes fail to improve utilization and efficiency, their main effect will be to increase total health care expenditures, as the American experience suggests. In both cases providers can supplement incomes through direct charges, but whereas for the uniform deductible and coinsurance schemes providers were restricted by a binding fee schedule, these differential charge schemes remove all direct pricing restrictions. Physicians would be freed from the constraint presently imposed in Ontario by negotiation with OHIP over fee schedules. Thus, compared to the present universal first-dollar coverage scheme these two forms of direct charges would likely lead to higher levels of health care costs.

It was argued that service repackaging, selective deinsurance, parallel systems, and incentives for self-care would all tend to encourage desirable shifts in levels and patterns of utilization. In addition, to the extent that such schemes not only alter utilization patterns but also place competitive pressures on providers through the creation of alternative forms of health care delivery, they will influence the efficiency with which care is delivered.

A natural division therefore seems to present itself between those schemes which have been most frequently proposed and implemented (all the uniform-charge schemes, plus MRM and extra-billing) and the more limited range

of less familiar, less well-defined alternatives. With the former group, we can now, so to speak, see the tunnel at the end of the light. The major weakness of the uniform-charge plans as cost control mechanisms is that uniformity allows no scope for price-sensitive consumers to affect provider market shares. In all cases, to the extent that initial deterrence effects do materialize, secondary utilization feedback effects will work in the opposite direction to maintain provider incomes. Furthermore, it will not necessarily be the marginally useful care which is deterred, because in most cases the services in question do not satisfy the product-characteristic/consumer-information precondition.

But differential charges do not guarantee that one will achieve success even with respect to the two system objectives – utilization patterns and efficiency. While differential charges appear to be a necessary condition, they are unfortunately not sufficient. It is not clear that such charges are an appropriate rationing device for many of the goods and services offered by the health care market. But as we emphasized earlier, if the mechanism of differential charges affecting consumer choice, in turn affecting provider efficiency, is deemed inappropriate, the case for direct charges to consumers on technical efficiency grounds collapses.

We have shown that in most cases this mechanism is inappropriate because of the peculiar nature of the market. Where consumers have a marked information deficiency compared to suppliers and/or where implicit or explicit structural rigidities (e.g. collusion, lack of advertising, etc.) exist, use of the differential charge nexus is clearly inappropriate. A few examples of situations in which the precondition of informed consumers could be satisfied and structural rigidities could be removed were offered. Among them were service repackaging such as the splitting of prescription drug costs into ingredient and dispensing components and introducing direct, ambience-related charges for the latter; selective deinsurance for particular inefficacious, or at least non-health-related, procedures; and incentives to self-care for certain conditions. There may be more. But even for some of these few potential areas the informational and operational requirements seem to border on the prohibitive.

In most cases where some potential seems to exist, what appears to be required is a concerted effort to increase the volume and analytic rigour of evaluation in the form of cost-benefit and cost-effectiveness analyses. Included in these, of course, is the randomized controlled trial process of identifying inefficacious services (for which the 'health benefit' is by definition zero). Only after unequivocal identification of inefficacious services, or confirmation of equal therapeutic outcomes (for parallel systems), should we

attempt to tackle the second problem, the dissemination of information to both patient and provider.

In summary, then, it should by now be clear that the scope for deployment of direct charges as a strategy for cost containment or efficiency enhancement is extremely limited. Where it was established that both product-characteristic/consumer-information and market-structure/conduct preconditions could theoretically be satisfied, provider opposition (either direct or through indirect feedback effects, both perhaps being more survival instinct than malicious opposition) is likely to impede successful implementation. The purported benefits of direct charges to patients either are absent or seem to accrue primarily to providers. By reason of the identity linking health costs and provider incomes, it seems fair to suggest, therefore, that indiscriminate introduction of such charges runs the not inconsiderable risk of creating a health cost snare from which it may be far more difficult to get disentangled than from the one in which some perceive us to be caught at present.

It is customary at this stage of a paper to offer a few insightful directions for the continued pursuit of knowledge at the margin. But the tone of this and preceding chapters has been predominantly negative, indicating our assessment that the scope for using direct charges to patients to control health care costs is small. It should not come as a surprise, then, that we find few proposals, old or new, in the 'direct charges area' worth further serious consideration.

In the present atmosphere of increasing attention to cost containment, it would be gratifying to be able to close with a one line solution to the 'problem'. There may, indeed, be such a solution, but we can be fairly confident in suggesting that its existence will have little, if anything, to do with charging patients. Only when the consumer has a chance to make informed choices among therapeutically equivalent service packages can we expect him to bear the financial consequences of such choices. In the present structure of health care delivery, most proposals for 'patient participation in health care financing' reduce to misguided or cynical efforts to tax the ill and/or to drive up the total cost of health care while shifting some of the burden out of government budgets. Abstract economic demonstrations of the social optimality of price-guided expressions of consumer sovereignty do not extend to the market offer, 'Your money or your life'.

References

- Aday, L.A. and R. Andersen (1975) *Access to Medical Care* (Ann Arbor: Health Administration Press)
- Anderson, W.M. (1952) 'Voluntary health and pension plans — can they meet the major needs?' *Proceedings* Institute of Public Administration of Canada
- Andreopoulos, S., ed. (1975) *National Health Insurance: Can We Learn From Canada?* (New York: John Wiley and Sons)
- Arrow, K.J. (1963) 'Uncertainty and the welfare economics of medical care'. *American Economic Review* 53, 941-73
- Babson, J.H. (1973) *Disease Costing* (Manchester: Manchester University Press)
- Babson, J.H. et al. (1975) 'Disease costing applied to measuring the cost of treating patients undergoing surgical removal of the gall bladder'. Paper presented at the University of Montreal, 29-31 January 1975
- Badgley, R. (forthcoming 1979) *Report of the Task Force to the Ontario Council of Health on User Charges* (Toronto: Ontario Council of Health)
- Barer, M.L. (1977) 'A methodology for derivation of marginal costs of hospital cases, and application to estimation of cost savings from community health centres'. Ph.D dissertation, University of British Columbia
- (1979) 'Community health centres and hospital expenditures in Ontario'. Mimeo. (Toronto: Ontario Economic Council)
- Barer, M.L. and R.G. Evans (1978) 'A hospital sector-specific deflator for time series analysis'. Mimeo.

- Barer, M.L. and P. Manga (1978) 'A benefit incidence and utilization analysis of the hospital insurance program in Ontario: methodology and some preliminary results'. Paper presented to the Canadian Economics Association, University of Western Ontario, 30 May 1978
- Bartlett, John (1955) *Bartlett's Familiar Quotations*, 13th ed. (Boston: Little, Brown and Co.)
- Beck, R.G. (1971) 'An analysis of the demand for physicians' services in Saskatchewan'. Ph.D dissertation, University of Alberta
- (1973) 'Economic class and access to physician services under public medical care insurance'. *International Journal of Health Services* 3, 341-55
- (1974) 'The effects of copayment on the poor'. *Journal of Human Resources* 9, 129-42
- (1976) 'Some dynamic effects of copayment on utilization of medical services in Saskatchewan'. In R.D. Fraser, ed., *Health Economics Symposium* (Kingston: Queen's University, Industrial Relations Centre) 98-110
- Beck, R.G. and J.M. Horne (1976) 'Economic class and risk avoidance: experience under public medical care insurance'. *Journal of Risk and Insurance* 43, 73-86
- (1978) 'An analytical overview of the Saskatchewan copayment experiment in the hospital and ambulatory care settings'. Report for the Ontario Council of Health. Mimeo.
- Benham, L. (1972) 'The effect of advertising on the price of eyeglasses'. *Journal of Law and Economics* 15, 337-52
- Benham, L. and A. Benham (1975) 'Regulating through the professions: a perspective on information control'. *Journal of Law and Economics* 18, 421-47
- Bennett, J. and J. Krasny (1977) 'Our health care stands tall but it doesn't come cheap'. *Financial Post*, 26 March, 7
- Blumberg, M. (1977) 'Changed provider prices as an incentive for improved health services'. (Oakland: Kaiser Foundation Health Plan)
- Boulet, J. and G. Grenier (1978) 'Health expenditures in Canada and the impact of demographic changes on future government health insurance program expenditures'. Discussion Paper No. 123 (Ottawa: Economic Council of Canada)
- Brian, E.W. and S.F. Gibbens (1974) 'California's Medi-Cal copayment experiment'. *Medical Care* 12 December, Supplement 1-56
- British Columbia (1975) *Report of the Children's Dental Health Research Project*. Report prepared for the Department of Health and the College of Dental Surgeons of B.C. (Victoria: Queen's Printer for B.C.)

- Bunker, J.P. (1970) 'A comparison of operations and surgeons in the United States and in England and Wales'. *New England Journal of Medicine* 282, 135-44
- Bunker, J.P. et al. (1977) *Costs, Risks and Benefits of Surgery* (New York: Oxford University Press)
- Cady, J.F. (1975) *Drugs on the Market* (Lexington: D.C. Heath)
- Canada (1963) *Voluntary Medical Insurance in Canada 1955-1961*. Memorandum Number 17 (Ottawa: Department of National Health and Welfare)
- Canada (1964) *Report of the Royal Commission on Health Services* (Ottawa: Queen's Printer)
- Canada (1970) *Task Force Reports on the Cost of Health Services in Canada* Vols 1-3 (Ottawa: Queen's Printer)
- Canada (1970-8) *Canadian Statistical Review* (Ottawa: Statistics Canada)
- Canada (1973a) *Expenditure on Personal Health Care in Canada 1960-1971* (Ottawa: Health and Welfare Canada)
- Canada (1973b) *National Health Expenditures in Canada, 1960-71* (Ottawa: Health and Welfare Canada)
- Canada (1975) *National Health Expenditures in Canada, 1960-73* (Ottawa: Health and Welfare Canada)
- Canada (1976) *Combines Investigations Act*, R.S.C. 314, s.1 (Ottawa: Queen's Printer)
- Canada (1978) *National Income and Expenditure Accounts, 2nd Quarter 1978* (Ottawa: Statistics Canada)
- Canada (forthcoming) *Report of the Task Force on the Periodic Health Examination* (Ottawa: Health and Welfare Canada)
- Cochrane, A.L. (1972) *Effectiveness and Efficiency: Random Reflections on Health Services* (London: Nuffield Provincial Hospitals Trust)
- Cohen, J.M. and J.J. Cohen, eds, (1960) *Penguin Dictionary of Quotations* (Middlesex, England: Penguin Books)
- Contandriopoulos, A.P. and J. Lance (1976) 'Un modele de Prevision de la Main d'Oeuvre Medicale'. Mimeo. Final report of NHG project 605-21-48, McGill University.
- Culyer, A.J. (1976) *Need and the National Health Service* (Totowa, N.J.: Rowman and Littlefield)
- (1978) *Measuring Health: Lessons for Ontario* Ontario Economic Council Research Study (Toronto: University of Toronto Press)
- Darsky, B.J. et al. (1958) *Comprehensive Medical Service Under Voluntary Health Insurance – A Study of Windsor Medical Services* (Cambridge, Massachusetts: Harvard University Press)

- Davies, G.N. (1973) 'Fluoride in the prevention of dental caries: a tentative cost-benefit analysis'. *British Dental Journal* [a series of six articles, May through October]
- Economic Council of Canada (1970) *Patterns of Growth* Seventh Annual Review (Ottawa: Queen's Printer)
- Eisner, R. and R.H. Strotz (1961) 'Flight insurance and the theory of choice'. *Journal of Political Economy* 69, 355-68
- Enterline, P.E. (1973) 'The distribution of medical services before and after 'free' medical care — the Quebec experience'. *New England Journal of Medicine* 289, 1174-8
- Enterline, P.E. et al. (1973) 'Effects of free medical care on medical practice — the Quebec experience'. *New England Journal of Medicine* 288, 1152-5
- Enthoven, A.C. (1978) 'Consumer choice health plan' (Parts I and II). *New England Journal of Medicine* 298, 650-8 and 709-20
- Evans, R.G. (1972) *Price Formation in the Market for Physician Services in Canada, 1957-1969*. Study for the Prices and Incomes Commission (Ottawa: Queen's Printer)
- (1974) 'Supplier-induced demand: some empirical evidence and implications'. In M. Perlman, ed., *The Economics of Health and Medical Care* (London: MacMillan)
- (1975) 'Beyond the medical marketplace: expenditure, utilization, and pricing of insured health care in Canada'. In S. Andreopoulos, ed., *National Health Insurance: Can We Learn from Canada?* (New York: John Wiley and Sons) 129-78
- Evans, R.G. and G.C. Robinson (1973) *An Evaluation of the Economic Implications of a Day Care Surgery Unit*. Final Report, N.H.W. Grant # 610-21-14 (Vancouver: University of British Columbia)
- Evans, R.G. et al (1975) *Economic Implications of an Ambulatory Diagnostic and Treatment Centre for Handicapped Children*. Final Report, N.H.W. Grant # 610-1029-21 (Vancouver: University of British Columbia)
- Evans, R.G. and M.F. Williamson (1978) *Extending Canadian Health Insurance: Options for Pharmacare and Denticare*. Ontario Economic Council Research Study (Toronto: University of Toronto Press)
- Evans, R.G. and A.D. Wolfson (1978) 'Moving the target to hit the bullet: generation of utilization by physicians in Canada'. Paper prepared for the National Bureau of Economic Research Conference on the Economics of Physician and Patient Behaviour, Stanford, 27-8 January 1978
- Feldstein, M. (1970) 'The rising price of physicians' services'. *Review of Economics and Statistics* 52, 121-33
- (1974) 'Econometric studies of health economics'. In M.D. Intriligator and

- D.A. Kendrick, eds, *Frontiers in Quantitative Economics* (Amsterdam: North-Holland) 377-447
- (1976) 'Seven principles of social insurance'. *Challenge* Nov.-Dec., 6-11
 - (1977) 'The high cost of hospitals — and what to do about it'. *The Public Interest* 48, 40-54
- Feldstein, M. and E. Allison (1972) 'Tax subsidies of private health insurance: distribution, revenue loss, and effects'. Discussion paper 237. Mimeo. Harvard Institute of Economic Research
- Feldstein, M. and B. Friedman (1974) 'Tax subsidies: the rational demand for insurance, and the health care crisis'. Discussion Paper 382. Mimeo. Harvard Institute of Economic Research
- Forester, D.J. and E.M. Schultz, eds, (1974) *International Workshop on Fluoride and Dental Caries Reductions* (Baltimore, Maryland: University of Maryland, School of Dentistry, Department of Pedodontics)
- French, Susan (1978) *The Cost-Containment Literature: Index and Summary*. Background Paper (Appendix D) for the *Report of the Select Committee on Health Care Financing and Costs* (Queen's Printer for Ontario)
- Friedman, E.A. et al. (1978) 'Pragmatic realities in uremia therapy'. *New England Journal of Medicine* 298, 368-71
- Fuchs, V. (1974) *Who Shall Live: Health, Economics and Social Choice* (New York: Basic Books)
- Fuchs, V. and M. Kramer (1972) *Determinants of Expenditure for Physicians' Services in the United States 1948-68*, study for National Centre for Health Services Research and Development, and National Bureau of Economic Research (Washington D.C.: Department of Health, Education and Welfare)
- Goldberg, L.G. and W. Greenberg, (1977) 'The effect of physician-controlled health insurance'. *Journal of Health Politics, Policy and Law* 2, 48-78
- Goldstein, G.S. and M.V. Pauly (1976) 'Group health insurance as a local public good'. In R. Rosett, ed., *The Role of Health Insurance in the Health Services Sector* (New York: National Bureau of Economic Research)
- Gross, John (1978) 'Present and future costs of health services to the aged in Ontario'. Paper presented at the Sixth Health Research Review Seminar, Ontario Economic Council, 11 July 1978
- Hall, C. (1966) 'Deductibles in health insurance: an evaluation'. *Journal of Risk and Insurance* 23, 253-63
- (1974) 'Impact of cost-sharing on consumer use of health services'. In S.

- Mushkin, ed., *Consumer Incentives for Health Care* (New York: Prodist) 333-61
- Hastings, J.E.F. et al. (1973) 'Prepaid group practice in Sault Ste Marie'. *Medical Care* 11, 91-103
- Havighurst, C. (1974) 'Speculations on the market's future in health care'. *Regulating Health Facilities Construction* (Washington D.C.: American Enterprise Institute)
- (1977) 'Controlling health care costs: strengthening the private sector's hand'. *Journal of Health Politics, Policy and Law* 1, 471-98
- Helms, J., J.P. Newhouse, and C.E. Phelps (1978) 'Copayments and demand for medical care: the California medical experience'. *Bell Journal of Economics* 9, 192-208
- Hoeckelman, R.A. (1975) 'What constitutes adequate well-baby care?' *Pediatrics* 55, 313-26
- Horne, J.M. (1978) 'Copayment and utilization of publicly insured hospital services in Saskatchewan: an empirical analysis'. Ph.D dissertation, Carleton University
- Huff, D. (1954) *How to Lie with Statistics* (New York: W.W. Norton and Co.)
- Illich, I. (1975) *Medical Nemesis* (Toronto: McClelland and Stewart)
- Irazuzta, J. (1979) 'A trend analysis of hospital utilization in Canada'. M.Sc. Thesis, Department of Health Administration, University of Toronto
- Keeler, E.B. et al. (1974) 'Deductibles and the demand for medical services — the theory of a consumer facing a variable price schedule under uncertainty'. Paper R-1514-OEO/NC (Santa Monica: Rand Corporation)
- (1977) 'The demand for supplementary health insurance, or do deductibles matter?' *Journal of Political Economy* 85, 789-801
- Kessel, R.A. (1958) 'Price discrimination in medicine'. *Journal of Law and Economics* 1, 20-53
- Korcok, M. (1978) 'Why all these Americans are coming to study Canada's health care system'. *Canadian Medical Association Journal* 118, 81-3
- Krizay, J. and A. Wilson (1974) *The Patient as Consumer: Health Care Financing in the United States* (Toronto: D.C. Heath)
- Kushman, J. and R. Scheffler (1975) 'Rationing professional services: some empirical evidence from dentistry'. Mimeo. University of California at Davis, Department of Agricultural Economics
- Lalonde, M. (1974) *A New Perspective on the Health of Canadians* (Ottawa: Health and Welfare Canada)
- Lay, C.M. (1978) 'Disease Costing in an Ambulatory Clinic: Disease and

- Physician Profiles and the Selection of Patients for Review'. Ph.D dissertation, Massachusetts Institute of Technology
- LeClair, M. (1975) 'The Canadian health care system'. In S. Andreopoulos, ed., *National Health Insurance: Can We Learn From Canada?* (Toronto: John Wiley and Sons)
- Lindblom, C.E. (1977) *Politics and Markets: The World's Political-Economic Systems* (New York: Basic Books)
- Manga, P. (1978) *The Income Distribution Effect of Medical Insurance in Ontario* (Toronto: Ontario Economic Council)
- Marmor, T.R. and E. Tenner (1977) 'National health insurance: Canada's path, America's choices'. *Challenge*, May-June, 3-21
- McKeown, T. (1976) *The Role of Medicine* (London: Nuffield Provincial Hospitals Trust)
- Moore, M. (1973) *How Much Price Competition?* (Montreal: McGill-Queen's University Press)
- Morgan, R.W. (1977) *Prospects for Preventive Medicine* (Toronto: Ontario Economic Council)
- Mueller, M.S. (1977) 'Private health insurance in 1975: coverage, enrolment, and financial experience'. *Social Security Bulletin* 40: 6, 3-21
- Mueller, M.S. and R.M. Gibson (1976) 'National health expenditures, fiscal year 1975'. *Social Security Bulletin* 39: 2, 3-20
- Mueller, M.S. and P.A. Piro (1976) 'Private health insurance in 1974: a review of coverage, enrolment and financial experience'. *Social Security Bulletin* 39: 3, 3-20
- Newhouse, J.P. (1976) 'Health care cost sharing and cost containment'. Paper p-5615 (Santa Monica: Rand Corporation)
- Newhouse, J.P. and C.E. Phelps (1974) 'On having your cake and eating it too: an analysis of estimated effects of insurance on demand for medical care'. R-1149-NC (Santa Monica: Rand Corporation)
- (1976) 'New estimates of price and income elasticities of medical care services'. In R. Rosett, ed., *The Role of Health Insurance in the Health Services Sector* (New York: National Bureau of Economic Research)
- Newhouse, J.P., C.E. Phelps, and W.B. Schwartz (1974) 'Policy options and the impact of national health insurance'. *New England Journal of Medicine* 290, 1345-59
- Newhouse, J.P., J.E. Rolf, B. Mori, and M. Murphy (1978) 'An estimate of the impact of deductibles on the demand for medical care services'. Memorandum R-1661-HEW (Santa Monica: Rand Corporation)
- Newhouse, J.P. and V. Taylor (1969) 'A new approach to hospital insurance'. (Santa Monica: Rand Corporation)

- Ontario (1975) *Ontario Statistics 1975* (Toronto: Ministry of Treasury, Economics and Intergovernmental Affairs)
- (1976) 'Financing health insurance in Ontario'. *Budget Paper B, Ontario Budget 1976* (Toronto: Ministry of Treasury, Economics and Intergovernmental Affairs)
 - (1977a) *Ontario Statistics 1977* (Toronto: Ministry of Treasury, Economics and Intergovernmental Affairs)
 - (1977b) *Report of the Joint Advisory Committee of the Government of Ontario and the Medical Association on Methods to Control Health Care Costs* (Taylor Report) (Toronto: Queen's Printer)
 - (n.d. [1978]) *OHIP Practitioner Care Statistics 1976-77 (pre-audit)* (Toronto: Ministry of Health)
 - (1978) *Testimony to the Select Committee of the Ontario Legislature on Health Care Financing and Costs* 10 July-15 Sept. 1978 (Toronto: Queen's Printer)
- Ontario Economic Council (1976) *Issues and Alternatives – Health* (Toronto: Ontario Economic Council)
- Perlman, M., ed., (1974) *The Economics of Health and Medical Care*. Proceedings of a Conference of the International Economics Association (London: Macmillan)
- Phelps, C.E. and J.P. Newhouse (1973) *Coinurance and the Demand for Medical Services*. Monograph R-964-OEO-NC (Santa Monica: Rand Corporation)
- Rennie, D. (1978) 'Home dialysis and the costs of uremia' *New England Journal of Medicine* 298, 399-400
- Rivard, J.Y. and A.P. Contandriopoulos (1977) 'Evolution du profil de pratique des médecins du Québec 1971 à 1973' (preliminary draft). Département de l'administration de santé, Université de Montréal
- Roemer, M.I. (1961) 'Bed supply and hospital utilization: a natural experiment'. *Hospitals J.A.H.A.*, 1 Nov., 36-42
- Roemer, M.I. et al. (1972) *Health Insurance Effects: Services, Expenditures and Attitudes under Three Types of Plans*. Bureau of Public Health Economics Research Series 16 (Ann Arbor, Mich.: University of Michigan, School of Public Health)
- (1975) 'Copayments for ambulatory care: penny wise and pound foolish'. *Medical Care* 13, 457-66
- Roos, L.L. et al. (1978) 'Assessing the impact of tonsillectomies'. *Medical Care* 16, 502-18
- Roos, N.P. (1977) 'Who should do the surgery? A comparison of T & A when performed by otolaryngologists, general surgeons and general practi-

- tioners'. Mimeo. Department of Social and Preventive Medicine, University of Manitoba
- Roos, N.P. et al. (1977a) 'A new audit procedure applied to an old question: is the frequency of T & A justified?' *Medical Care* 15, 1-19
- (1977b) 'Elective surgical rates: do high rates mean lower surgical standards?' *New England Journal of Medicine* 297, 360-5
- Sackett, D.L. (1972) 'The family physician and the periodic health examination'. *Canadian Family Physician* 18, 61-5
- Salkever, D. (n.d. [1975]) 'The problem of access to medical care: a consumer demand analysis'. Mimeo. Working Paper in Economics 11. Department of Political Economy, Johns Hopkins University
- Scitovsky, A.A. and N. McCall (1977) 'Coinsurance and the demand for physician services: four years later'. *Social Security Bulletin* 40: 5, 19-27
- Scitovsky, A.A. and N.M. Snyder (1972) 'Effect of coinsurance on use of physician services'. *Social Security Bulletin* 35: 6, 3-19
- Shillington, C.H. (1972) *The Road to Medicare in Canada* (Toronto: Del Graphics)
- Silverman, M. and P.R. Lee (1974) *Pills, Profits and Politics* (Berkeley: University of California Press)
- Stanbury, W. and G. Reschenthaler (1976) 'Oligopoly and conscious parallelism: theory, policy and the Canadian cases'. Mimeo. Faculty of Commerce and Business Administration, University of British Columbia
- Stoddart, G.L. (1975) 'An episodic approach to the demand for medical care'. Ph.D dissertation, University of British Columbia
- Taylor, M.G. (1957) *Financial Aspects of Health Insurance* (Toronto: Canadian Tax Foundation)
- Trout, K. (1963) 'A review of several features of health insurance plans'. *Journal of Health Insurance Administration* 15, 169-91
- United States (1977) *Health – United States – 1976-1977* Washington D.C.: Department of Health, Education, and Welfare)
- (1978) *A National Profile of Catastrophic Illness*. National Center for Health Services Research and Development, Research Summary Series, DHEW publication (PHS) 78-3201 (Washington D.C.)
- Vayda, E. (1973) 'A comparison of surgical rates in Canada and in England and Wales'. *New England Journal of Medicine* 289, 1224-9
- Walker, H.D. (1971) *Market Power and Price Levels in the Ethical Drug Industry* (Bloomington: Indiana University Press)
- Walton, R.J. et al. (1976) 'Cervical cancer screening programs' (parts I-IV). *Canadian Medical Association Journal* 114, 1003-33

- Wolfson, A.D. (1975) 'The supply of physicians' services in Ontario'. Ph.D dissertation, Harvard University
- Wolfson, A.D. and A.J. Solari (1976) 'Research report on the results of the patient utilization study'. Unpublished report, Ontario Ministry of Health
- Wolfson, A.D. and C.J. Tuohy (forthcoming, 1979) *Opting Out of OHIP* (Toronto: Ontario Economic Council)
- Yankauer, A. (1973) 'Child health supervision: is it worth it?' *Pediatrics* 52, 272-7

Ontario Economic Council Occasional Papers

- 1 Basic Skills at School and Work: the study of Albertown an Ontario community**
O. Hall and R. A. Carlton
- 2 Prospects for Preventive Medicine: a catalogue**
R. W. Morgan
- 3 Input-Output Analyses of Fiscal Policy in Ontario**
R. W. Boadway, A. A. Kubursi and J. M. Treddenick
Edited by J. Bossons
- 4 Educational Problems in Ontario and Some Policy Options**
J. A. Buttrick
- 5 The Market for New Housing in the Metropolitan Toronto Area**
R. A. Muller
- 6 The Income Distribution Effect of Medical Insurance in Ontario**
P. Manga
- 7 Who Benefits from the Ontario University System:
a benefit-cost analysis by income groups**
O. Mehmet
- 8 Financial Markets and Foreign Ownership**
J. C. Pattison
- 9 The Administrative Cost of Income Security Programs:
Ontario and Canada**
M. Mendelson
- 10 Controlling Health Care Costs by Direct Charges to Patients:
Snare or Delusion?**
M. L. Barer, R. G. Evans and G. L. Stoddart

These publications are available from the Ontario Government Bookstore,
880 Bay Street, Toronto, Ontario M7A 1N8
